

1961

# Index selection as influenced by errors of parameter estimation

Dewey Lynn Harris  
*Iowa State University*

Follow this and additional works at: <https://lib.dr.iastate.edu/rtd>



Part of the [Genetics Commons](#)

---

## Recommended Citation

Harris, Dewey Lynn, "Index selection as influenced by errors of parameter estimation " (1961). *Retrospective Theses and Dissertations*. 2434.  
<https://lib.dr.iastate.edu/rtd/2434>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact [digirep@iastate.edu](mailto:digirep@iastate.edu).

This dissertation has been 61-3036  
microfilmed exactly as received

HARRIS, Dewey Lynn, 1933-  
INDEX SELECTION AS INFLUENCED BY  
ERRORS OF PARAMETER ESTIMATION.

Iowa State University of Science and Technology  
Ph.D., 1961  
Biology - Genetics

University Microfilms, Inc., Ann Arbor, Michigan

INDEX SELECTION AS INFLUENCED BY ERRORS  
OF PARAMETER ESTIMATION

by

Dewey Lynn Harris

A Dissertation Submitted to the  
Graduate Faculty in Partial Fulfillment of  
The Requirements for the Degree of  
DOCTOR OF PHILOSOPHY

Major Subject: Animal Breeding

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University  
Of Science and Technology  
Ames, Iowa

1961

## TABLE OF CONTENTS

	Page
INTRODUCTION	1
THEORY OF SELECTION INDEXES	3
ESTIMATION OF GENETIC AND PHENOTYPIC PARAMETERS	24
CALCULATED INDEXES	36
PROGRESS FROM INDEX SELECTION	38
VARIANCES AND COVARIANCES OF ESTIMATES	59
SIMULATION OF ESTIMATION FROM MULTI-SIRE SAMPLES WITH TWO TRAITS	72
RESULTS	81
DISCUSSION	101
SUMMARY	118
BIBLIOGRAPHY	121
ACKNOWLEDGEMENTS	128
APPENDIX A	129

## INTRODUCTION

To clarify the principles of making maximum genetic improvement by selection is a primary goal of research in animal and plant breeding. In the majority of breeding operations, several traits have economic importance. Even in cases where only one trait is economically important, other traits can be useful aids to selection due to their relationship with the economically important trait. To make maximum genetic change, a certain optimum emphasis must be placed upon each trait.

Smith (1936) presented a method for determining an index for the selection of superior varieties. Hazel (1943) presented an equivalent method for developing an index for the selection of genetically superior individuals. The index, in both cases, is a linear function of certain observed characteristics of each individual or variety in the population within which selection is to be practiced. The coefficients for this index may be calculated from certain genetic and phenotypic parameters.

In practice the genetic and phenotypic parameters necessary to determine the coefficients for an index are not known exactly and estimates of these parameters must be used in making the calculations. Recent studies have indicated that, with the amount of data usually available in large animal species, estimates of genetic parameters have especially large

sampling errors. It seems clear that inaccuracies of estimation would reduce genetic progress from index selection, as compared to what would be attainable if the true parameters were known, but the magnitude of this reduction has not been studied.

The purpose of the present study was to consider the inaccuracies in the estimation of parameters and to determine their influence upon the progress and upon the estimation of progress which results from selection by means of indexes calculated from estimates based upon varying amounts of data. Of primary interest was the situation where the estimates are derived from analyses of variance and covariance among and within sire progeny groups, although some of the formulae shall be applicable for other types of estimation procedures. Both mathematical and empirical techniques were used for this investigation.

## THEORY OF SELECTION INDEXES

The theory of simultaneous selection for several traits in an effort to make maximum genetic improvement was first developed by Smith (1936). He used the concept of discriminant functions as presented by Fisher (1936) to obtain a technique for selecting among varieties of wheat on the basis of yield data from plots of each of the varieties. This technique is actually a special application of multiple regression techniques. The symbolism and terminology used below differ considerably from that used by Smith, but the theoretical aspects are the same.

Conceptually, a genotypic value for net worth exists for each variety. Smith considered the genotypic net worth of a variety to be a linear function of the genotypic values for each economically important trait, each weighted by the value of one unit change in that trait. Thus, by definition,

$$H = a_1G_1 + a_2G_2 + \cdots + a_nG_n = \sum_{i=1}^n a_iG_i,$$

where  $H$  is the genotypic net worth of a variety,  
 $a_i$  is the relative value of one unit change in the  $i^{\text{th}}$  character,  
 $G_i$  is the average genotypic value for the  $i^{\text{th}}$  character of individuals in the variety,  
 $n$  is the number of traits being considered,

and  $\sum_{i=1}^n$  indicates the summation of terms having subscript values of  $i$  from 1 up to and including  $n$ .

The traits of interest are numbered from 1 up to and including  $n$  for the purpose of symbolism. The term "genotypic value" as used here represents the contribution to the phenotype or observed characteristic due to all the genes possessed by an individual. Thus "average genotypic value" will be the mean value for all individuals in the variety, an average value being necessary since these values will vary except for the unusual case where genotypes of all the individuals in the variety are alike.

The goal of a selection program among varieties should be directed toward producing maximum improvement in the genotypic net worth by eliminating from the population as many as possible of those varieties which have the lowest genotypic net worth. Since the genotypes of a variety will not be known exactly, neither will the genotypic net worth be known. Thus, direct selection for genotypic net worth cannot be practiced. However, selection may be based upon an index,  $I$ , determined from the observable characteristics of each variety. In Smith's study the observed characteristics were measurements of important characters of several plots of each variety.

The average genotypic net worth of a group of varieties selected on the basis of their index values would have an expectation of



$$E(\bar{H}) = \mu_H + B_{HI}(\bar{I}_S - \mu_I),$$

where  $E(\bar{H})$  is the expectation of the average H in the selected group,

$\mu_H$  is the population mean for H before selection was practiced,

$B_{HI}$  is the linear regression coefficient of H on I,

$\bar{I}_S$  is the average index value in the selected group,

and  $\mu_I$  is the population mean for I.

This equation is based on the supposition that the regression of H on I is truly linear. The reasonableness of this assumption will be examined later.

The superiority in genotypic net worth of a selected group as compared with the original population before selection is  $B_{HI}(\bar{I}_S - \mu_I)$ . This may be advantageously expressed as  $B_{HI} \sigma_I \bar{i}_S$  where  $\bar{i}_S$  is the mean of the selected group when variable I is coded into standard measure as

$$i = \frac{I - \mu_I}{\sigma_I},$$

and  $\sigma_I$  is the standard deviation of index values. Hence,  $\bar{i}_S$  is equal to  $\frac{\bar{I}_S - \mu_I}{\sigma_I}$ . When truncation selection is performed

and I is normally distributed,  $\bar{i}_S$  will equal  $z/p$ , where  $z$  is

the height of the ordinate of the theoretical unit normal distribution of  $i$  at the point of truncation with fraction  $p$  of the population above that point. This interesting relationship follows from the fact that the integral of the values under the unit normal curve above a specified point is equal to the height of the ordinate at that point. This may be proven using the usual calculus notation as follows:

$$\int_k^{\infty} i \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2} i^2} di = \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2} k^2} = z, \text{ say,}$$

where  $k$  is the point of truncation of  $i$  values. The mean value for individuals above the truncation point is the ratio of the integral of values above that point to the fraction of individuals having such values. Thus,

$$\bar{i}_s = \frac{z}{p}, \text{ where } p = \int_k^{\infty} \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2} i^2} di.$$

Numerical values of  $z$  and  $p$  for different  $k$  are given in convenient form by Anderson and Bancroft (1952) and in several other sets of tables. When  $k$  is the point of truncation for  $i$  values, the corresponding point for index ( $I$ ) values will be  $\mu_I + k\sigma_I$ .

A convenient form for the index is

$$I = b_1 X_1 + b_2 X_2 + \cdots + b_n X_n = \sum_{i=1}^n b_i X_i$$

where the  $X_i$ 's are the observed characters that give the best

indication of the  $G_i$ 's of the variety. This linear form is usually used due to the resulting mathematical simplifications. The coefficients in the index, the  $b_i$ 's, must be chosen so that  $B_{HI} \sigma_I$  will be maximized, and thus  $B_{HI} \sigma_I \bar{i}_s$ , the expected genetic gain from selection, will be maximized, since  $\bar{i}_s$  will be a constant.

The relationship between the  $X_i$ 's and the  $G_i$ 's is assumed to be

$$X_i = G_i + E_i$$

where the  $G_i$  and  $E_i$  are independent elements which go together additively to produce  $X_i$ . With the previous definition of  $G_i$  as the average genotypic value for the  $i^{\text{th}}$  trait of individuals in the variety,  $E_i$  will be the deviation of the observed character ( $X_i$ ) from the  $G_i$  due to environmental influences upon each particular plot plus a deviation, probably small, due to the difference in average genotypic value of the individuals in the plot under consideration from the average genotypic value in the entire variety. A further assumption that there is no covariance between the  $G_i$  value for one trait and the  $E_i$  for that trait or for any other trait is necessary for the development of the theory to follow. Thus, the variance of  $X_i$  is equal to the variance of  $G_i$  plus the variance of  $E_i$ .

If the symbol  $G_{ij}$  is chosen to represent the covariance

between  $G_i$  and  $G_j$  and the symbol  $P_{ij}$  is to represent the covariance between  $X_i$  and  $X_j$ , it is found that

$$V(I) = V \left( \sum_{i=1}^n b_i X_i \right) = \sum_{i=1}^n \sum_{j=1}^n b_i b_j P_{ij}$$

and

$$\begin{aligned} \text{Cov}(HI) &= \text{Cov} \left[ \left( \sum_i a_i G_i \right), \left( \sum_j b_j X_j \right) \right] \\ &= \sum_{i=1}^n \sum_{j=1}^n a_i b_j \text{Cov}(G_i, X_j) = \\ &\quad \sum_{i=1}^n \sum_{j=1}^n a_i b_j G_{ij}. \end{aligned}$$

The symbol  $V( )$  represents the variance of the quantity in the parentheses, and similarly  $\text{Cov}( )$  symbolizes the covariance of the two quantities in the parentheses. Of course,  $P_{ii} = V(X_i)$  and  $G_{ii} = V(G_i)$ . Therefore,

$$B_{HI} \sigma_I = \frac{\sum_{ij} a_i b_j G_{ij}}{\left( \sum_{ij} b_i b_j P_{ij} \right)^{1/2}}.$$

The  $b_i$  values which maximize  $B_{HI} \sigma_I$  are those which satisfy the  $n$  simultaneous equations resulting from successively setting equal to zero the first partial derivative of  $B_{HI} \sigma_I$  with respect to each of the  $b_i$ 's. The  $t^{\text{th}}$  equation of the set of  $n$  equations with  $n$  unknowns is obtained as follows:

$\frac{\partial B_{HI} \sigma_I}{\partial b_t}$  = first partial derivative of  $B_{HI} \sigma_I$  with

$$\text{respect to } b_t = \frac{\sum_i a_i G_{it}}{(\sum_{ij} b_i b_j P_{ij})^{1/2}}$$

$$- \frac{(\sum_{ij} a_i b_j G_{ij})(\sum_i b_i P_{it})}{(\sum_{ij} b_i b_j P_{ij})^{3/2}} = 0$$

or

$$\frac{\sum_i b_i P_{it}}{\sum_{ij} b_i b_j P_{ij}} = \frac{\sum_i a_i G_{it}}{\sum_{ij} a_i b_j G_{ij}}.$$

Smith (1936) then took as his equations the following:

$$\sum_i b_i P_{it} = \frac{1}{B_{HI}} \sum_i a_i G_{it} \text{ for } t = 1, 2, \dots, n.$$

However, these  $n$  equations are satisfied when the  $n$  equations

$$\sum_i b_i P_{it} = \sum_i a_i G_{it}$$

for  $t = 1, 2, \dots, n$  are satisfied, for then

$$\begin{aligned} \frac{\sum_i a_i G_{it}}{\sum_{ij} a_i b_j G_{ij}} - \frac{\sum_i b_i P_{it}}{\sum_{ij} b_i b_j P_{ij}} &= \frac{\sum_i a_i G_{it}}{\sum_j b_j \sum_i a_i G_{ij}} \\ &- \frac{\sum_i a_i G_{it}}{\sum_j b_j \sum_i a_i G_{ij}} = 0. \end{aligned}$$

Therefore, it seems unnecessary to have the  $\frac{1}{B_{HI}}$  in the equations. Later it will be shown that  $B_{HI}$  is equal to unity when  $b_i$  values are obtained from these equations.

Hazel (1943) developed an index similar to Smith's for use in swine selection. Their techniques of development differed in that Hazel maximized  $R_{IH}$ , the correlation between the index values and genotypic net worth of individuals in the population, because superiority in the average genotypic net worth,

$$E(\bar{H} - \mu_H) = B_{HI} \sigma_I \bar{i}_s = R_{IH} \sigma_H \bar{I}_s ,$$

and  $\sigma_H$ , the standard deviation of genotypic net worth values, is constant for a particular population. However, the two maximization procedures yield equivalent results because the above relationship between  $B_{HI} \sigma_I$  and  $R_{IH}$  exists. Hazel expressed the equations for determining the  $b_i$  values in terms of correlations rather than covariances, the  $t^{\text{th}}$  of the set of  $n$  equations is

$$\sum_i B_i r_{X_i X_t} = r_{X_t H} ,$$

where  $B_i = b_i \frac{p_{ii}^{\frac{1}{2}}}{\sigma_H} ,$

$r_{X_i X_t}$  is the correlation between  $X_i$  and  $X_t$ ,

and  $r_{X_t H}$  is the correlation between  $X_t$  and  $H$ .

The equivalence between these two forms of the  $t^{\text{th}}$  equation may be seen by examining the equation,

$$\sum_i b_i P_{it} = \sum_i a_i G_{it} .$$

Multiplying left and right hand members of the equation by  $\frac{1}{P_{tt}^{\frac{1}{2}} \sigma_H}$  and multiplying the  $i^{\text{th}}$  term of the left hand member by  $\frac{P_{ii}^{\frac{1}{2}}}{P_{ii}^{\frac{1}{2}}}$  = 1 yields

$$\sum_i b_i \frac{P_{it}}{P_{ii}^{\frac{1}{2}} P_{tt}^{\frac{1}{2}}} \frac{P_{ii}^{\frac{1}{2}}}{\sigma_H} = \sum_i a_i \frac{G_{it}}{P_{tt}^{\frac{1}{2}} \sigma_H}$$

or

$$\sum_i B_i r_{X_i X_t} = \frac{\text{Cov}(X_t, H)}{P_{tt}^{\frac{1}{2}} \sigma_H} = r_{X_t H} .$$

The  $r_{X_t H}$  occurring in these equations may be obtained by noting the following relations

$$\begin{aligned} r_{X_t H} &= \frac{\text{Cov}(X_t, H)}{P_{tt}^{\frac{1}{2}} \sigma_H} = \frac{\sum_i a_i \text{Cov}(X_t, G_i)}{P_{tt}^{\frac{1}{2}} \sigma_H} \frac{G_{ii}^{\frac{1}{2}}}{G_{ii}^{\frac{1}{2}}} \\ &= \sum_i d_i r_{X_t G_i}, \text{ where } d_i = a_i \frac{G_{ii}^{\frac{1}{2}}}{\sigma_H} . \end{aligned}$$

Thus, there are two forms of the simultaneous equations for obtaining the coefficients for an index. These forms are as follows:

$$\sum_i b_i P_{it} = \sum_i a_i G_{it} \text{ for } t = 1, 2, \dots, n,$$

or

$$\sum_i B_i r_{X_1 X_t} = r_{X_t H} \text{ for } t = 1, 2, \dots, n.$$

The second is the coded form of the first, each variable in the second being measured on a scale on which its own standard deviation is unity. Either form may be used, but the latter may be preferable if the data to be used for the calculations are in terms of correlations. The former may be more advantageous for computational purposes if the phenotypic and genotypic variances and covariances can be readily obtained from the data. If a solution exists, the set of equations may be solved to yield the  $b_i$  values in an index by any of the methods of solving simultaneous equations--by successive elimination, by the ratio of determinants technique (Cramer's rule), or by matrix inversion.

Since the operations of matrix algebra [see, for example, Aitken, (1954) and Browne, (1958)] will be used in later portions of this paper, it should be pointed out that the first of the above forms for the equations may be written in matrix notation as follows:

$$Pb = Ga$$

where  $P$  is the  $n \times n$  matrix ( $n$  rows and  $n$  columns) of  $P_{ij}$  values,



$b$  is the  $n \times 1$  matrix (column vector of  $n$  values) of  $b_i$  values,

$G$  is the  $n \times n$  matrix of  $G_{ij}$  values,

and  $a$  is the  $n \times 1$  matrix of  $a_i$  values.

Thus,  $b = P^{-1} Ga$  where  $P^{-1}$  symbolizes the inverse matrix of the  $P$  matrix. This matrix form has been used in numerous papers; see for example, Kempthorne and Nordskog (1959).

It should be pointed out that selection of individuals as studied by Hazel (1943) differs somewhat from variety selection although the algebraic treatment is equivalent, as is shown above. As discussed by Fisher (1918) and Wright (1921), the portion of the genotype for a particular trait that tends to be transmitted from parent to offspring is in effect the sum of the average effects (within that population) of the individual genes which influence that characteristic. The value for this portion is usually termed the additive genetic value, and the variance of such a value is known as the additive genetic variance. So for individual selection the symbolism used up to this point is redefined as follows:

$G_i$  is the additive genetic value of a particular individual for the  $i^{\text{th}}$  trait,

$E_i$  is the deviation of the phenotype for the  $i^{\text{th}}$  trait from the additive genetic value due to allelic and nonallelic interactions and due to environmental influences,

$G_{ii}$  is the additive genetic variance for the  $i^{\text{th}}$  trait,

and  $G_{ij}$  is the covariance between the additive genetic values for the  $i^{\text{th}}$  and  $j^{\text{th}}$  traits.

This redefinition allows the use of the previously presented theory for individual selection as well as selection among varieties.

Hazel (1943) also pointed out that for individual selection the  $X_i$  values could be one record on an individual's performance, the average of several records on the individual's performance, or observations on parents, collateral relatives, or offspring of the individual. He went on to show what  $r_{G_i X_j}$  would be for each of several cases involving observations on relatives. Of course, the economic value,  $a_i$ , for the traits of relatives would be zero if the additive genetic value for that trait were also included in the equation for  $H$ .

$$\text{Since } \sum_i b_i P_{it} = \sum_i a_i G_{it},$$

$$\text{Cov}(I, H) = \sum_{i=1}^n \sum_{j=1}^n a_i b_j G_{ij} = \sum_{i=1}^n \sum_{j=1}^n b_i b_j P_{ij} = V(I).$$

Therefore,

$$R_{IH} = \frac{\text{Cov}(I, H)}{\sigma_I \sigma_H} = \frac{\sigma_I}{\sigma_H},$$

$$B_{HI} = \frac{\text{Cov}(I, H)}{V(I)} = \frac{V(I)}{V(I)} = 1,$$

$$\text{and } B_{IH} = \frac{\text{Cov}(I, H)}{V(H)} = \frac{V(I)}{V(H)} = R_{IH}^2.$$

The second of these relationships leads to an alteration of the formula for the superiority in H of a selected group.

That is,

$$B_{HI} \sigma_I \bar{i}_s = \sigma_I \bar{i}_s \text{ since } B_{HI} = 1.$$

Therefore, the genetic improvement is  $(\bar{i}_s - \mu_I)$ , which is the selection differential of index values.

Utilizing regression theory, Morley (1950) showed that the expected genetic improvement in the  $t^{\text{th}}$  character may be determined by obtaining the regression of  $G_t$  on I. Since

$$\text{Cov}(G_t, I) = \text{Cov}(G_t, \sum_i b_i X_i) = \sum_i b_i G_{it},$$

$$B_{G_t I} = \frac{\sum_i b_i G_{it}}{\sigma_I^2}.$$

So the change in  $G_t$  which may be expected to accompany a given change in I when selection is based upon I is

$$B_{G_t I} \sigma_I \bar{i}_s = \bar{i}_s \frac{\sum_i b_i G_{it}}{\sigma_I}.$$

The expected phenotypic selection differential may be similarly obtained.

$$\text{Since } \text{Cov}(X_t, I) = \text{Cov}(X_t, \sum_i b_i X_i) = \sum_i b_i P_{it}, \text{ so that}$$

$$B_{X_t I} = \frac{\sum_i b_i P_{it}}{\sigma_I^2},$$

the phenotypic selection differential for the  $t^{\text{th}}$  trait is

$$\bar{i}_s \frac{\sum_i b_i P_{it}}{\sigma_I} = \bar{i}_s \frac{\sum_i a_i G_{it}}{\sigma_I}$$

Manning (1956) presented an alternative approach to the application of selection index theory. In his problem of selecting for yield improvement in cotton, the phenotype for economic net worth,  $X_W$ , was directly observable. With a relationship the same as that for phenotypic and additive genetic values for other traits,

$$X_W = H + E_W,$$

the  $t^{\text{th}}$  of the set of  $n$  simultaneous equations would be

$$\sum_i b_i P_{it} = G_{tW}.$$

Phenotypic net worth,  $X_W$ , might, but need not necessarily, be included among the  $X_i$  values comprising the index. It would in most cases be desirable to include  $X_W$  in the index unless it was difficult to assess or unless all of its components were included in the other traits used. This approach bypasses the necessity of defining the economic weights, the  $a_i$ 's, assigned to each trait, although some function of the component traits may be necessary to obtain the  $X_W$  values. However, it is not necessary that this latter function be

linear. The covariance,  $G_{tW}$ , automatically determines the linear relationship between the additive genetic value for net worth,  $H$ , and the additive genetic value for the  $t^{\text{th}}$  trait,  $G_t$ . In those cases in which the phenotypic net worth of individuals can be assessed, this alternative approach avoids the sometimes difficult task of specifying the  $a_i$  values.

Lindholm and Stonaker (1957) presented a technique for obtaining the  $a_i$  values which might not be reliable in certain circumstances. They used as  $a_i$  values the regression coefficients from the multiple regression equation for  $X_W$  on the  $X_i$  for  $i=1, 2, \dots, n$ . As used in index theory, the  $a_i$  values are the multiple regression coefficients of  $H$  on the  $G_i$  values. In matrix notation the phenotypic multiple regression coefficients are obtained from the relations

$$Pa = P_W \text{ or } a = P^{-1} P_W$$

where  $a$  is the  $n \times 1$  matrix of  $a_i$  values,  
 $P$  is the  $n \times n$  matrix of  $P_{ij}$  values,  
 $P_W$  is the  $n \times 1$  matrix of  $P_{iW}$  values,  
and  $P^{-1}$  is the inverse matrix of the  $P$  matrix.

Since  $G_{tW} = \sum_i a_i G_{it}$  when  $H = \sum_i a_i G_i$ ,  $G_W = Ga$  or  $a = G^{-1} G_W$

is the relationship which defines the  $a_i$  values,

where  $G_W$  is the  $n \times 1$  matrix of  $G_{iW}$  values,  
 $G$  is the  $n \times n$  matrix of  $G_{ij}$  values,

and  $G^{-1}$  is the inverse matrix of the  $G$  matrix.

So,  $b = P^{-1} Ga$ , the  $n \times 1$  matrix of  $b_i$  values, is equal to

$$P^{-1} GG^{-1} G_W = P^{-1} G_W$$

when  $a = G^{-1} G_W$  and is equal to  $P^{-1} GP^{-1} P_W$

when  $a = P^{-1} P_W$

as in Lindholm and Stonaker's paper. These two results seem to be different except when  $P^{-1} P_W = G^{-1} G_W$ . This equality will result when

$$X_W = \sum_i a_i X_i \text{ and } H = \sum_i a_i G_i$$

are the true relationships. For then

$$P_{iW} = \sum_j a_j \overline{P_{ij}} \text{ and } G_{iW} = \sum_j a_j G_{ij},$$

or in matrix notation,

$$P_W = Pa \text{ and } G_W = Ga$$

are the true relationships and either set of equations should yield the same  $a_i$  values. When the true relationships are not linear, the approach used by Manning seems preferable.

Manning's approach does not avoid the restrictions of linear relationships, but avoids the confusion usually associated with specifying the  $a_i$  values. However, with the limited amount of data available to Lindholm and Stonaker, estimates of the genetic covariances necessary for Manning's approach would probably not have been accurate.

According to Tabler and Touchberry (1955), Henderson

(1951) presented an alternative approach to calculation of the coefficients for a selection index. This approach simplifies the calculation of several indexes involving different sets of  $a_i$  values. An index,  $I_i$ , which is of the form

$$I_i = \sum_j b_{ij} X_j$$

where the  $b_{ij}$  values are those which maximize  $B_{G_i I_i} \sigma_{I_i}$  may be determined for each trait. The equations for obtaining the  $b_{ij}$  values are

$$\sum_j b_{ij} P_{jt} = G_{it} \text{ for } t = 1, 2, \dots, n.$$

Such an index,  $I_i$ , may be formed for each trait, that is, for  $i = 1, 2, \dots, n$ . Then the over-all index will be

$$I = \sum_i a_i I_i = \sum_j b_j X_j \text{ as before.}$$

From this it is seen that

$$b_j = \sum_i a_i b_{ij}.$$

This procedure makes possible the calculation of coefficients for different indexes with different sets of  $a_i$  values from the relation,  $I = \sum_i a_i I_i$ , with much less effort than completely recalculating each set of coefficients. It is easily seen that this technique yields the same results as before.

Since

$$\sum_j b_{ij} P_{jt} = G_{it} \text{ for } t = 1, 2, \dots, n$$

and for  $i = 1, 2, \dots, n$  in matrix notation,

$$PB = G, \text{ or } B = P^{-1} G$$

where B is the  $n \times n$  matrix of  $b_{ij}$  values. Also, in matrix notation the equation

$$b_j = \sum_i a_i b_{ij} \text{ for } j = 1, 2, \dots, n \text{ becomes } b = Ba.$$

Thus  $b = P^{-1} Ga$  as before.

Kempthorne and Nordskog (1959) developed theory for restricted selection indexes; that is, indexes upon which some constraint has been imposed. For example, an index was developed which would hold egg size constant while correlated traits were being improved.

Cochran (1951) reviewed the theory of selection indexes and developed formulae for the optimum culling level at each stage when selection is carried out in two stages. He also considered the effect of discarding variates from an index when their importance seems small.

Manning (1956) and Hanson and Johnson (1957) considered the effect of interactions of genotype and environment upon the development of a selection index. Under such circumstances each "environment" will be associated with a different set of parameters. Manning's procedure was to calculate a selection index from the yield data each year and use that index for that year's selection. Hanson and Johnson noted the desirability of pooling data from different experiments, i.e., different locations and/or years, so as to minimize sampling



errors. They thereby developed theory which leads to an iterative process of pooling the data from different sources. This process maximizes the average genetic advance for all sources of data.

At this point a review of the assumptions involved in the preceding development seems to be in order. These assumptions were as follows:

- 1)  $X_i = G_i + E_i$  for all  $i$ ,
- 2)  $\text{Cov}(G_i, E_j) = 0$  for all  $i$  and  $j$ , and
- 3) the regression of  $H$  on any linear function of the  $X_i$  is truly linear.

These assumptions are common to much of statistical genetic theory and are used in most applications of this theory. The first one means that the phenotypic value of the  $i^{\text{th}}$  character is the sum of the genotypic value or additive genetic value and the effects of environment upon that character; that is, the two types of effects are combined additively. When this is not true, it is sometimes possible to transform the data to obtain such a relationship and then to use the transformed variables in the calculations. For example, when the two factors are combined multiplicatively, i.e.,

$$X_i = G_i \times E_i$$

a logarithmic transformation yields an additive relationship,

$$\log X_i = \log G_i + \log E_i.$$

However, the true relationship is rarely known exactly and

may be complex; it is thereby difficult to obtain the proper transformation, and the most appropriate simple (usually linear) "model" is the expedient used in such cases.

The second assumption requires that there is no correlation (or covariance) between the genotypic values for any particular trait and the deviations due to environmental influences for that trait or for any other trait. This assumption will not be wholly valid when there is a tendency (probably human) for the environmental influences to be allotted according to genetic merit of the individuals or varieties. This tendency should be avoided in any breeding operation.

When individual selection is being considered,  $G_i$  is the additive genetic value rather than the total genotypic value, and the dominance and epistatic deviations are included in the  $E_i$  value. Because of the definition of additive genetic value, these deviations will combine additively with it and will be uncorrelated with  $G_i$ .

The last assumption of linear regression of  $H$  on any linear function of the  $X_i$  will be satisfied if the two variables are distributed together in a bivariate normal distribution. Since both of these are linear functions of biological measures, and since many biological measures have been found to approach normality in their distribution, this assumption will be valid the majority of cases. Even if the

individual elements of the linear functions are not normally distributed, a linear function tends toward normality as the number of variables in the function increases.

The assumptions involved in the theory of selection indexes appear reasonable for most cases. However, these suppositions should be evaluated critically for each application of the theory, to assure validity and reliability of the results for the particular application.

## ESTIMATION OF GENETIC AND PHENOTYPIC PARAMETERS

The theory presented in the previous section was based entirely upon the true parameters of the population which was to be selected. However, in the application of this theory, the true parameters will not be known and estimates of these parameters will have to be used in the equations.

Throughout this paper estimates of a parameter will be denoted by a circumflex ( $\hat{\phantom{x}}$ ) placed over the symbol for the parameter. For example,  $\hat{P}_{ij}$  denotes an estimate of the parameter  $P_{ij}$ .

Because the parameters for the case of variety selection and for the case of individual selection are defined differently, estimation of the parameters for the two cases is discussed separately. For variety selection Smith (1936) proposed that the estimates be obtained from the mean squares calculated in analyzing the variance in a randomized block experiment as follows:

<u>Sources of Variation</u>	<u>Degrees of Freedom</u>	<u>Mean Square or Product</u>	<u>Expected Mean Square or Product</u>
Varieties	$v-1$	$v_{ij}$	$E(v_{ij}) = E_{ij} + bG_{ij}$
Error	$(v-1)(b-1)$	$e_{ij}$	$E(e_{ij}) = E_{ij}$
Total (excluding blocks)	$b(v-1)$	$t_{ij}$	$E(t_{ij}) = E_{ij} + G_{ij}$

where  $v$  is the number of varieties in the experiment,  
 $b$  is the number of blocks,

$v_{ij}$  is the variety mean square ( $i=j$ ) or mean product ( $i \neq j$ ) for the  $i^{\text{th}}$  and  $j^{\text{th}}$  characters,

$e_{ij}$  is the error mean square or mean product,

$t_{ij}$  is the total (excluding blocks) mean square or mean product,

$E()$  is the expectation or population mean value of the variable in the parentheses,

and  $E_{ij}$  is the error variance or covariance component on a plot basis.

When selection is to be based upon  $k$  replications of each variety while the experiment was based on  $b$  replicates, the estimates are

$$\hat{G}_{ij} = \frac{v_{ij} - e_{ij}}{b} \quad \text{with } E(\hat{G}_{ij}) = G_{ij}$$

and

$$\hat{P}_{ij}(k) = \hat{G}_{ij} + \frac{e_{ij}}{k} = \frac{v_{ij}}{b} + e_{ij} \left( \frac{1}{k} - \frac{1}{b} \right)$$

with

$$E[\hat{P}_{ij}(k)] = G_{ij} + \frac{E_{ij}}{k}$$

where  $P_{ij}(k)$  indicates the phenotypic variance or covariance when each phenotype is the mean observation from  $k$  plots.

This estimation procedure results from the fact that with uncorrelated environmental effects upon the plots in different blocks, the component of variance for varieties is the measure of the variation among the average genotypic values for the

varieties. Similarly the error mean square is a measure of the variation on a plot basis due to environmental influences plus a small portion due to genetic variation within a variety. Of course, when  $k = b$ ,

$$\hat{P}_{ij}(b) = \frac{v_{ij}}{b} \quad \text{and} \quad E[\hat{P}_{ij}(b)] = G_{ij} + \frac{E_{ij}}{b}.$$

The individual plots are considered to be of the same size for the experiment and for selection. When this is not the case, the situation is complicated considerably. Of course, experimental designs other than a randomized block design could be used to obtain estimates of the parameters.

When selection is to be among individuals, the estimation procedure is somewhat different. The phenotypic parameters may be estimated from the observed relationships in a sample from the population. For example, the sample covariance between the measurements of the  $i^{\text{th}}$  trait and those for the  $j^{\text{th}}$  trait may be the estimate,  $\hat{P}_{ij}$ , of the population phenotypic covariance,  $P_{ij}$ .

Estimation of the additive genetic variances and covariances is more difficult since each individual has a unique genotype, confounded with environmental influences. Even when repeated observations can be made on an individual's phenotype, the effects of some of these environmental influences persist over some or all of the periods of observations and, hence, do not tend to be averaged out and cannot be

separated from the genotypic contributions. The usual procedure for avoiding this difficulty has been to estimate additive genetic variances and covariances from the covariances between measurements on individuals which have some genes in common but which have, as much as is possible, independent environmental contributions.

Extending earlier studies, Fisher (1918) and Wright (1921) determined the correlations between relatives in terms of the additive genetic variances and phenotypic variances as well as variances due to allelic and non-allelic interactions. These results have been extended in later studies by these and other workers. Lush (1940) presented techniques for estimating heritability values from certain regressions and correlations among traits measured in relatives, heritability values being defined as the ratio of additive genetic variance to total phenotypic variance.

Hazel (1943) presented techniques for estimating genetic correlations, the correlations between additive genetic values for two traits. All of these techniques are based on the fact that for a genetically influenced trait, related animals tend to resemble each other more than unrelated animals. Kempthorne (1957) presented the general equation to represent the covariance between relatives for all types of relationships as follows:

$$\text{Cov}(w^{X_i}, z^{X_i}) = \sum_{\substack{r+s=n \\ r=0, s=0 \\ \text{except } 0, 0}} (2r_{WZ})^r (U_{WZ})^s \sigma_{ArDs}^2$$

where W and Z represent individuals related by a certain type of relationship,

$w^{X_i}$  and  $z^{X_i}$  are the measures of the  $i^{\text{th}}$  trait in individuals W and Z, respectively,

$r_{WZ}$  is the probability that a random gene at a certain locus of W is identical by descent to a random gene of Z at this same locus (i.e.,  $r_{WZ}$  is Malécot's (1948) "coefficient de parenté" and  $2r_{WZ}$  is the numerator of Wright's (1921) "coefficient of relationship".),

$U_{WZ}$  is the probability that both genes at a particular locus of W are identical by descent to both genes at this locus for Z,

and  $\sigma_{ArDs}^2$  is the component of variance due to all interactions of  $r$  additive effects and  $s$  dominance deviations. (This general term includes  $\sigma_A^2$ , the additive genetic variance when  $r=1$  and  $s=0$  and  $\sigma_D^2$ , the dominance variance when  $r=0$  and  $s=1$ . All other values are due to non-allelic or epistatic interactions.)



The summation goes up to all values of  $r$  and  $s$  such that  $r + s$  equals the total number of loci segregating. This formula is derived on the supposition of random mating in the population and ignores the effects of linkage and selection. Also, the assumption is made that there is no covariance between the environmental influences upon  $W$  and those upon  $Z$ . Kempthorne also presented the values of  $2r_{WZ}$  and  $U_{WZ}$  for certain types of relationships which are as follows:

<u>Relationship</u>	<u><math>2r_{WZ}</math></u>	<u><math>U_{WZ}</math></u>
Identical twins	1	1
Full sibs	1/2	1/4
Parent-offspring	1/2	0
Parent- $k^{\text{th}}$ generation descendant	$(1/2)^k$	0
Uncle-nephew	1/4	0
Cousins	1/8	0
Double first cousins	1/4	1/16

From these values it is seen that the covariance between a parent and its  $k^{\text{th}}$  generation descendant is

$$(1/2)^k \sigma_A^2 + (1/2)^{2k} \sigma_{AA}^2 + (1/2)^{3k} \sigma_{AAA}^2 + (1/2)^{4k} \sigma_{AAAA}^2 + \dots$$

It is difficult to evaluate the magnitude of the variance components for epistatic interactions, but due to the smallness of the coefficients of these components, especially for those representing high order interactions, any effects of selection

upon future generations will be primarily through increases in the average additive genetic values of individuals in intervening generations. This conclusion is reached by noting that as  $k$ , the number of intervening generations, increases the coefficients of the interaction variance components become smaller at a faster rate than does the coefficient of  $\sigma_A^2$ , the additive genetic variance component. This to a large extent justifies the use of the simplified model,

$$X_i = G_i + E_i,$$

when considering the effects of selection among individuals as related to genetic improvement. This model denotes the situation as if allelic and non-allelic gene interactions were of the same nature as temporary environmental influences which do not affect individuals in future generations. This representation is reasonably close for most situations, although not necessarily exactly correct for any particular situation. With this model the covariance between relatives for the  $i^{\text{th}}$  trait becomes simply  $(2r_{WZ})G_{ii}$ , and the covariance between the  $i^{\text{th}}$  trait in an individual and the  $j^{\text{th}}$  trait in its relative becomes  $(2r_{WZ})G_{ij}$ . Hence, an estimate,  $\hat{G}_{ij}$ , of the additive genetic covariance may be obtained by multiplying the calculated covariance between  ${}_W X_i$  and  ${}_Z X_j$  by  $\frac{1}{2r_{WZ}}$ .

When using this procedure for estimating additive genetic variances and covariances, one should be constantly aware of the assumptions involved in its use and of the possible

biases which could be introduced into the estimates by (1) the effects of selection, (2) the inclusion of some of the epistatic variation, and (3) environmental covariation between the relatives. Other possible pitfalls include nonadditivity of  $G_i$  and  $E_i$ , covariation between  $G_i$  and  $E_i$ , sex linked effects, hereditary maternal effects influencing the traits, and nonrandomness of mating.

Lush (1940, 1948) discussed in some detail the possible influences of these inadequacies of the model upon estimation of heritability values. For the purpose of index construction, primary interest is upon the estimates of variances and covariances. In general, selection will decrease variation, and epistatic gene actions will bias estimates of additive genetic variance upward. Environmental covariation between relatives will also bias the the estimates of additive genetic variation upward if the covariation is positive. The writings of Lush give a more detailed discussion of these influences. Therefore, they will not be considered in further detail here.

The covariances between relatives may be estimated from a sample of pairs of relatives in the population, such as daughter-dam pairs, or from the analysis of variance and covariance of groups of related individuals, such as paternal half sibs. The analysis of variance or covariance table for paternal-half-sibs is as follows:

<u>Sources of Variation</u>	<u>Degrees of Freedom</u>	<u>Mean Square or Product</u>	<u>Expected Mean Square or Product</u>
Among sire groups	$s-1$	$s_{ij}$	$E(s_{ij}) = \sigma_{ije} + m_0 \sigma_{ijs}$
Within sire groups	$\sum_k (m_k - 1) = M - s$	$w_{ij}$	$E(w_{ij}) = \sigma_{ije}$
Total	$\sum_k m_k - 1 = M - 1$	$t_{ij}$	$E(t_{ij}) = \sigma_{ije} + \frac{m_0(s-1)}{M-1} \sigma_{ijs}$

where  $s$  is the number of sire-groups in the sample and the sires are numbered from 1 up to and including  $n$ ,  $m_k$  is the number of offspring in the group for the  $k^{\text{th}}$  sire,

$M = \sum_{k=1}^s m_k$  is the total number of offspring in the experiment,

$s_{ij}$  is the mean square ( $i=j$ ) or mean product ( $i \neq j$ ) for variation among sire groups,

$w_{ij}$  is the mean square or mean product for variation within sire groups,

$t_{ij}$  is the mean square or mean product for the total variation among groups and within groups,

$\sigma_{ije}$  is the component of variance ( $i=j$ ) or covariance ( $i \neq j$ ) for the within group variation,

$\sigma_{ijs}$  is the component of variance for the variation among sire groups,

and  $m_0$  is equal to

$$\frac{1}{s-1} \left[ \sum_k m_k - \frac{\sum_k m_k^2}{\sum_k m_k} \right] = \frac{M}{s} - \frac{V(m_k)}{M}$$

where

$$V(m_k) = \frac{\sum_k m_k^2 - \frac{1}{s} (\sum_k m_k)^2}{s-1}.$$

The usual assumptions for analyses of variance or covariance with a hierarchical classification are involved in this table.

Based upon the general equation for the covariance between relatives as presented above, the genetic interpretations of the variance and covariance components as pointed out by Hazel and Terrill (1945) are

$$\sigma_{ije} = 3/4 G_{ij} + E_{ij}$$

and 
$$\sigma_{ijs} = 1/4 G_{ij}$$

where 
$$E_{ij} = P_{ij} - G_{ij}.$$

Thus,  $\hat{G}_{ij}$ , the estimate of  $G_{ij}$ , would be  $4 \left[ \frac{s_{ij} - w_{ij}}{m_0} \right]$ .

The assumptions involved here are that (1) the sires and dams were a random sample from the population, (2) the dams were allotted to the sires at random, and (3) there was no selection among the progeny. Each dam can have only one offspring.

This analysis of variance or covariance table leads to two possible estimators of  $P_{ij}$ , the phenotypic variance or covariance. These are  $t_{ij}$ , the total mean square, or the sum of the estimates of the among and within sire groups

components,

$$\hat{\sigma}_{ije} + \hat{\sigma}_{ijs} = w_{ij} + \frac{s_{ij} - w_{ij}}{m_0} = \frac{m_0 - 1}{m_0} w_{ij} + \frac{1}{m_0} s_{ij}.$$

In terms of the additive genetic and environmental variances or covariances,

$$\begin{aligned} E(t_{ij}) &= P_{ij} - 1/4 G_{ij} + \frac{m_0(s-1)}{4(M-1)} G_{ij} \\ &= P_{ij} - \frac{m_0(s-1) - (M-1)}{4(M-1)} G_{ij} \\ &= P_{ij} - \frac{\sum_k m_k^2 - M}{4M(M-1)} G_{ij}, \end{aligned}$$

since  $P_{ij} = G_{ij} + E_{ij}$ ,

and  $E(\hat{\sigma}_{ije} + \hat{\sigma}_{ijs}) = G_{ij} + E_{ij}$ .

From these expectations it is seen that the sum of the estimates of the components is an unbiased estimator, while the total mean square or product is negatively biased by the amount

$$\frac{\sum_k m_k^2 - M}{4M(M-1)} G_{ij} = \frac{\sum_k m_k^2 - M}{M(M-1)} \sigma_{ijs}.$$

The second moment of an estimator about the true value is usually considered to be an effective measure of the accuracy of this estimator. For the case of unbiased estimators, the second moment about the true value is the variance, while for

biased estimators the second moment is the variance plus the square of the bias. Evaluation of the second moments around the true value for these two estimators of  $P_{ij}$  will be deferred to a later section where the necessary formulae are developed.

## CALCULATED INDEXES

The theory of selection indexes, using various estimation procedures, has been used to develop several indexes for the selection of genetically superior individuals, varieties, or lines. Besides Hazel (1943), Warren and Dickerson (1952) and Sutherland (1958) calculated indexes for selecting swine. For sheep, selection indexes were developed by Hazel and Terrill (1946), Winters et al. (1946), Morley (1950), Rae (1950), Ercanbrack (1952), Karam, et al. (1953), Sidwell (1954), Felts (1958), Karam (1959), and Givens, et al. (1960).

Selection indexes for dairy cattle were presented by Legates and Lush (1950), Harvey and Lush (1952), Tabler and Touchberry (1955, 1959), and Young, et al. (1960). Evans (1956) and Lindholm and Stonaker (1957) calculated indexes for selection of beef cattle.

Indexes for poultry selection have been developed by Panse (1946), Lerner, et al. (1947), Krueger, et al. (1952), Wyatt (1954), Ghostley (1955), Farnsworth (1956), Merritt (1956), Hogsett and Nordskog (1958), Nordskog and Hill (1958), Yamada (1958), and Crittenden (1958).

Other indexes have been presented by Smith (1936) for wheat, Manning (1956) for cotton, and Nagai, et al. (1955) for mice.

The techniques of estimation and the volume of data used vary considerably among these indexes. For example, Harvey



and Lush (1952) used daughter-dam analysis on data involving 8,464 cows, while Sutherland (1958) used variance component analysis on measurements from only 937 individuals.

## PROGRESS FROM INDEX SELECTION

When estimates of genetic and phenotypic parameters are substituted for the true values in the equations to compute the  $b_i$  values which are intended to maximize genetic improvement, estimates of these  $b_i$  values, symbolized by  $\hat{b}_i$ , will be obtained. The accuracy of the  $\hat{b}_i$  values will depend upon the accuracy of the estimates of the parameters which are used in the equations. Since these  $\hat{b}_i$  values are not likely to be exactly the true  $b_i$  values, the expected genetic improvement for an index with coefficients calculated in this manner will be somewhat less than the improvement from an index based upon the true parameters, if these were known. Also, the estimated genetic progress as calculated from the estimates will differ from the true expected genetic progress for a particular calculated index.

When estimates of the genetic and phenotypic parameters are substituted in the equations for obtaining the coefficients for an index, the equations may be represented by

$$\sum_i \hat{b}_i \hat{P}_{it} = \hat{G}_{tW} \text{ for } t = 1, 2, \dots, n,$$

or in matrix notation

$$\hat{P} \hat{b} = \hat{G}_W$$

where  $\hat{P}$  is the  $n \times n$  matrix of  $\hat{P}_{ij}$  values,

$\hat{b}$  is the  $n \times 1$  matrix of  $\hat{b}_i$  values which are estimates of the optimum coefficients, the  $b_i$ 's,

and  $\hat{G}_W$  is the  $n \times 1$  matrix of  $\hat{G}_{jW}$  values.

The expected progress in improving  $H$  was previously presented as  $\bar{I}_S(\sum_i b_i G_{iW})^{\frac{1}{2}}$ , which is equal to  $\bar{I}_S(\sum_{ij} a_i b_j G_{ij})^{\frac{1}{2}}$  when  $H = \sum_i a_i G_i$ , and is equal to  $\bar{I}_S(\sum_{ij} b_i b_j P_{ij})^{\frac{1}{2}}$  in either case.

When errors of estimation of parameters are considered, the expected progress for a particular calculated index is

$$\bar{I}_S B_{HI} \hat{\sigma}_{\hat{I}} = \bar{I}_S \frac{\sum_i \hat{b}_i G_{iW}}{(\sum_{ij} \hat{b}_i \hat{b}_j P_{ij})^{\frac{1}{2}}},$$

where

$$\hat{I} = \sum_i \hat{b}_i X_i$$

represents the values for a particular calculated index. This expression for expected progress is derived by noting that the calculated index values have a variance,

$$V_C(\sum_i \hat{b}_i X_i) = \sum_{ij} \hat{b}_i \hat{b}_j P_{ij},$$

and a covariance with  $H$ ,

$$\text{Cov}_C(\sum_i \hat{b}_i X_i, H) = \sum_i \hat{b}_i G_{iW}.$$

When

$$H = \sum_i a_i G_i,$$

$$\text{Cov}_C(\hat{I}, H) = \sum_{ij} a_i \hat{b}_i G_{ij},$$

and

$$\bar{I}_S B_{HI} \hat{\sigma}_{\hat{I}} = \bar{I}_S \frac{\sum_{ij} a_i \hat{b}_j G_{ij}}{(\sum_{ij} \hat{b}_i \hat{b}_j P_{ij})^{\frac{1}{2}}}.$$

In matrix notation,

$$V_C(\hat{I}) = \hat{b}' P \hat{b},$$

where  $\hat{b}'$  is the  $1 \times n$  matrix which is the transpose of the  $\hat{b}$  matrix and

$$\text{Cov}_C(\hat{I}, H) = G'_W \hat{b}.$$

Thus 
$$\bar{I}_S B_{HI} \sigma_{\hat{I}} = \bar{I}_S \frac{G'_W \hat{b}}{(\hat{b}' P \hat{b})^{1/2}}.$$

When  $H = \sum_i a_i G_i$ ,  $G_W = Ga$  and 
$$\bar{I}_S B_{HI} \sigma_{\hat{I}} = \bar{I}_S \frac{a' G \hat{b}}{(\hat{b}' P \hat{b})^{1/2}}.$$

It should be noted that  $V_C(\hat{I})$  and  $\text{Cov}_C(\hat{I}, H)$  as presented here are the conditional variances and covariances for a particular set of  $\hat{b}_i$  values. Thus the subscript  $c$  as used here and later denotes that the subscripted function is a conditional function; that is, the particular function is evaluated with the  $\hat{b}_i$  values being considered as constants.

Thus the  $\bar{I}_S B_{HI} \sigma_{\hat{I}}$  term represents the genetic improvement expected when a particular calculated index,  $\hat{I}$ , is used as the basis for selection.

Bartlett (1939) discussed the theory of selection indexes or discriminant functions and the standard errors of the discriminant function coefficients. Nanda (1949a) extended the work of Bartlett and developed equations for the standard errors of the discriminant function coefficients when the estimation procedure results from a randomized block experiment as used by Smith. Nanda also presented equations for the standard errors of genetic advance when this type of estimation procedure is used. However, examination of the method of

derivation shows that the standard errors of genetic advance as presented by Nanda are actually the standard errors of estimates of genetic advance and not a measure of the variation in progress expected from different calculated indexes as the terminology used seems to imply. As will be shown later in this paper, the standard errors of estimates of genetic progress are not closely related to the standard deviations of expected progress values for calculated indexes. This work does point out the importance of including many varieties in the experiment.

In a later paper, Nanda (1949b) attempted to compare progress from index selection to selection where the phenotypes are weighted by their economic values. However, due to an error in the derivation of the formula used, the results are invalidated although the general conclusions reached seem plausible.

Hazel and Lush (1942) compared progress from index selection with the progress from two alternate selection procedures--tandem selection, selecting only one trait each generation and independent culling levels, the truncation culling of individuals having low merit in one of the traits of interest regardless of the individual's merit in the other traits. They studied the special case where all traits of interest were independent and the products of economic value, heritability, and phenotypic standard deviation are the same

for all traits; i.e.,  $a_i G_{ii}/P_{ii}^{1/2}$  is constant for all  $i$ . Recently, Young (1959) extended this work to include correlated traits with differing economic values, phenotypic variances, and heritability values. These studies lead to the conclusion that index selection will always result in as much improvement and in most cases greater improvement than the other two selection procedures. Young expresses concern that in some cases the difficulties and costs of index selection may offset its advantages over the simpler alternatives. However, neither of these studies considered the disturbing influences of inaccuracies of estimation.

Tallis (1960) developed equations for the variances and covariances of the index coefficients ( $b_i$  values in the symbolism used here) and the variance of predicted genetic gains. However, to the present author certain aspects of his development seem inadequate and the equations presented do not seem satisfactory for the purposes of the present study.

Hanson and Johnson (1957) pointed out that the correlation between the true index,  $I$ , and a calculated index,  $\hat{I}$ , may be denoted by

$$r_{I\hat{I}} = \frac{b' Pb}{(b' Pb)^{1/2} (\hat{b}' P\hat{b})^{1/2}}$$

since  $\text{Cov}_C(I, \hat{I}) = \text{Cov}(\sum_i b_i X_i, \sum_i \hat{b}_i X_i) = b' P\hat{b}$ ,

$$V(I) = b' Pb, \text{ and } V_C(\hat{I}) = \hat{b}' P\hat{b}.$$

Symbolizing  $\bar{i}_s B_{HI} \sigma_I = \bar{i}_s (b' G_W)^{\frac{1}{2}} = \bar{i}_s (b' Pb)^{\frac{1}{2}}$  by  $\Delta H$  and

$$\bar{i}_s B_{HI} \hat{\sigma}_I = \bar{i}_s \frac{G'_W \hat{b}}{(\hat{b}' Pb)^{\frac{1}{2}}} = \bar{i}_s \frac{b' P \hat{b}}{(\hat{b}' Pb)^{\frac{1}{2}}} \text{ by } \Delta H',$$

it is seen that

$$r_{II} \hat{=} \frac{\Delta H'}{\Delta H}.$$

This means that the correlation between true index values in a population and the corresponding values for a particular calculated index is equal to the ratio of progress expected from the calculated index to that expected from the true index.

Since a correlation coefficient must have a value between minus one and plus one,  $-1 \leq \frac{\Delta H'}{\Delta H} \leq 1$ , and thus  $-\Delta H \leq \Delta H' \leq \Delta H$ . This inequality is taken as proof of the statement that progress from an index based upon estimates must be less than or equal to the maximum progress attainable if the true parameters were known. The inequality also shows that the lower limit upon the expected progress from a calculated index is minus  $\Delta H$ .

As discussed by Kendall and Stuart (1958), use of the first few terms of a Taylor's expansion of a complex function leads to approximate equations for the expectation and variance of this complex function. Let  $g(x_1, x_2, \dots, x_k)$  be a function of the variables,  $x_1, x_2, \dots, x_k$ , and the  $x_i$  have means,  $m_i$ , and finite variances and be differentiable at

$x_i = m_i$ . Then the expansion of  $g$  is

$$g(x_1, x_2, \dots, x_k) = g(m_1, m_2, \dots, m_k) \\ + \sum_{i=1}^k \frac{\partial g}{\partial m_i} \Delta x_i + \frac{1}{2} \sum_{i,j=1}^k \frac{\partial^2 g}{\partial m_i \partial m_j} \Delta x_i \Delta x_j + O(\Delta x)^3$$

where  $\frac{\partial g}{\partial m_i}$  indicates the first partial derivative of  $g$  with respect to  $x_i$  evaluated at  $x_i = m_i$  for all values of  $i$ ,

$\Delta x_i$  is  $x_i - m_i$ ,

and  $O(\Delta x)^3$  indicates that the remaining terms are of the order  $(\Delta x)^3$ .

For the first moment,

$$E(g) \approx g(m_1, \dots, m_k) + \frac{1}{2} \sum_{i,j=1}^k \frac{\partial^2 g}{\partial m_i \partial m_j} \text{Cov}(x_i, x_j)$$

where the symbol  $\approx$  means "is approximately equal to" and

$\frac{\partial^2 g}{\partial m_i \partial m_j}$  indicates the second partial derivative of  $g$ , first with respect to  $x_i$ , then with respect to  $x_j$ , evaluated at  $x_i = m_i$  for all  $i$ . Also to the order of  $(\Delta x)^3$ ,

$$V(g) \approx \sum_{i,j=1}^k \frac{\partial g}{\partial m_i} \frac{\partial g}{\partial m_j} \text{Cov}(x_i, x_j) .$$

In a similar manner it may be shown that for two functions of random variables, say  $g$  and  $h$

$$\text{Cov}(g, h) \approx \sum_{i,j=1}^k \frac{\partial g}{\partial m_i} \frac{\partial h}{\partial m_j} \text{Cov}(x_i, x_j) .$$



The accuracy of these approximations depends upon the smallness of the high order terms of the expansion which are ignored. Application of these formulae leads to the general results

$$E\left(\frac{x}{y}\right) \cong \frac{E(x)}{E(y)} \left[ 1 + \frac{V(y)}{E^2(y)} - \frac{\text{Cov}(x, y)}{E(x) E(y)} \right],$$

$$V\left(\frac{x}{y}\right) \cong \frac{E^2(x)}{E^2(y)} \left[ \frac{V(x)}{E^2(x)} + \frac{V(y)}{E^2(y)} - 2 \frac{\text{Cov}(x, y)}{E(x) E(y)} \right],$$

$$\text{Cov}\left(x, \frac{y}{z}\right) \cong \frac{\text{Cov}(x, y)}{E(z)} - \frac{E(y) \text{Cov}(x, z)}{E^2(z)},$$

$$E(xy) = E(x) E(y) + \text{Cov}(x, y) \quad [\text{This is exact.}],$$

$$V(xy) \cong E^2(y) V(x) + E^2(x) V(y) + 2E(x) E(y) \text{Cov}(x, y),$$

$$E(x^{\frac{1}{2}}) \cong E^{\frac{1}{2}}(x) \left[ 1 - \frac{V(x)}{8E^2(x)} \right],$$

$$V(x^{\frac{1}{2}}) \cong \frac{V(x)}{4E(x)},$$

$$\text{Cov}(x, y^{\frac{1}{2}}) \cong \frac{\frac{1}{2} \text{Cov}(x, y)}{E^{\frac{1}{2}}(y)}$$

$$\text{and } \text{Cov}(x^{\frac{1}{2}}, y^{\frac{1}{2}}) \cong \frac{\text{Cov}(x, y)}{4E^{\frac{1}{2}}(x) E^{\frac{1}{2}}(y)}.$$

Utilizing these approximations, it is found that

$$E[\Delta H'] = \bar{i}_S E \left[ \frac{\text{Cov}_C(\hat{I}, H)}{V_C^{\frac{1}{2}}(\hat{I})} \right]$$

$$\begin{aligned}
& \approx \bar{I}_S \frac{E[\text{Cov}_C(\hat{I}, H)]}{E[V_C^{\frac{1}{2}}(\hat{I})]} \left[ 1 + \frac{V[V_C^{\frac{1}{2}}(\hat{I})]}{E^2[V_C^{\frac{1}{2}}(\hat{I})]} - \frac{\text{Cov}[\text{Cov}_C(\hat{I}, H), V_C^{\frac{1}{2}}(\hat{I})]}{E[\text{Cov}_C(\hat{I}, H)] E[V_C^{\frac{1}{2}}(\hat{I})]} \right] \\
& \approx \bar{I}_S \frac{E[\text{Cov}_C(\hat{I}, H)]}{E^{\frac{1}{2}}[V_C(\hat{I})] \left[ 1 - \frac{V_C[V(\hat{I})]}{8E^2[V_C(\hat{I})]} \right]} \left\{ 1 + \frac{V[V(\hat{I})]}{4E^2[V_C(\hat{I})] \left[ 1 - \frac{V[V_C(\hat{I})]}{8E^2[V_C(\hat{I})]} \right]^2} \right. \\
& \quad \left. - \frac{\text{Cov}[\text{Cov}_C(\hat{I}, H), V_C(\hat{I})]}{2E[\text{Cov}_C(\hat{I}, H)] E[V_C(\hat{I})] \left[ 1 - \frac{V[V_C(\hat{I})]}{8E^2[V_C(\hat{I})]} \right]} \right\}
\end{aligned}$$

In this and in future equations,  $V_C( )$  and  $\text{Cov}_C( )$  indicate the conditional variances and covariances, respectively, of the quantities in the parentheses for each calculated index, while  $E[ ]$ ,  $V[ ]$ , and  $\text{Cov}[ ]$  indicate the expectation, variance, and covariance, respectively, of the values in the brackets over the population of calculated indexes obtained from repeated estimations of the population parameters. Also,

$$\begin{aligned}
V[\Delta H'] &= \bar{I}_S^2 V \left[ \frac{\text{Cov}_C(\hat{I}, H)}{V_C^{\frac{1}{2}}(\hat{I})} \right] \\
&\approx \bar{I}_S^2 \frac{E^2[\text{Cov}_C(\hat{I}, H)]}{E^2[V_C^{\frac{1}{2}}(\hat{I})]} \left[ \frac{V[\text{Cov}_C(\hat{I}, H)]}{E^2[\text{Cov}_C(\hat{I}, H)]} + \frac{V[V_C^{\frac{1}{2}}(\hat{I})]}{E^2[V_C^{\frac{1}{2}}(\hat{I})]} \right. \\
& \quad \left. - 2 \frac{\text{Cov}[\text{Cov}_C(\hat{I}, H), V_C^{\frac{1}{2}}(\hat{I})]}{E[\text{Cov}_C(\hat{I}, H)] E[V_C^{\frac{1}{2}}(\hat{I})]} \right]
\end{aligned}$$

$$\begin{aligned}
& \left\{ \frac{E^2[\text{Cov}_C(\hat{I}, H)]}{E[V_C(\hat{I})] \left[ 1 - \frac{V[V_C(\hat{I})]}{8E^2[V_C(\hat{I})]} \right]^2} \frac{V[\text{Cov}_C(\hat{I}, H)]}{E^2[\text{Cov}_C(\hat{I}, H)]} \right. \\
& + \frac{V[V_C(\hat{I})]}{4E^2[V_C(\hat{I})] \left[ 1 - \frac{V[V_C(\hat{I})]}{8E^2[V_C(\hat{I})]} \right]^2} \\
& \left. - \frac{\text{Cov}[\text{Cov}_C(\hat{I}, H), V_C(\hat{I})]}{E[\text{Cov}_C(\hat{I}, H)] E[V_C(\hat{I})] \left[ 1 - \frac{V[V_C(\hat{I})]}{8E^2[V_C(\hat{I})]} \right]} \right\}
\end{aligned}$$

Denoting the matrix of errors of estimation of the phenotypic parameters,  $(\hat{P} - P)$  by  $\Delta p$ , the relation

$$\hat{P} = P + \Delta p = P(I_n + P^{-1} \Delta p),$$

where  $I_n$  is the identity matrix of size  $n \times n$ , is obtained. Similarly,

$$\hat{G}_W = G_W + \Delta g_W$$

where  $\Delta g_W$  is the matrix of errors of estimation of genetic parameters. When all the elements of the  $\hat{P}$  matrix are unbiased estimates of the corresponding elements of the  $P$  matrix; i.e.,  $E(\Delta p) = \underline{0}$  where  $\underline{0}$  is a matrix with all elements equal to zero,

$$\begin{aligned}
\hat{P}^{-1} &= [P(I_n + P^{-1} \Delta p)]^{-1} = [I_n + P^{-1} \Delta p]^{-1} P^{-1} \\
&\cong [I_n - P^{-1} \Delta p + P^{-1} \Delta p P^{-1} \Delta p] P^{-1}.
\end{aligned}$$

$$\begin{aligned}
\text{Then } \text{Cov}_C(\hat{I}, H) &= G_W' \hat{b} = G_W' \hat{P}^{-1} \hat{G}_W \cong G_W' [I_n - P^{-1} \Delta p \\
&\quad + P^{-1} \Delta p P^{-1} \Delta p] P^{-1} [G_W + \Delta g_W] \\
&\cong G_W' P^{-1} G_W - G_W' P^{-1} \Delta p P^{-1} G_W + G_W' P^{-1} \Delta p P^{-1} \Delta p P^{-1} G_W \\
&\quad + G_W' P^{-1} \Delta g_W - G_W' P^{-1} \Delta p P^{-1} \Delta g_W \\
&= b' G_W - b' \Delta p b + b' \Delta p P^{-1} \Delta p b + b' \Delta g_W - b' \Delta p P^{-1} \Delta g_W.
\end{aligned}$$

When the  $\hat{G}_{ij}$  values are also unbiased estimates,

$$\begin{aligned}
E[\text{Cov}_C(\hat{I}, H)] &\cong b' G_W + b' E[\Delta p P^{-1} \Delta p] b \\
&\quad - b' E[\Delta p P^{-1} \Delta g_W] \\
&= \text{Cov}(I, H) + \sum_{ijk} b_i P_{jk}^{-1} b_m \text{Cov}[\hat{P}_{ij}, \hat{P}_{km}] \\
&\quad - \sum_{ijk} b_i P_{jk}^{-1} \text{Cov}[\hat{P}_{ij}, \hat{G}_{kW}]
\end{aligned}$$

where  $P_{ij}^{-1}$  is the element in the  $ij$  position of the  $P^{-1}$  matrix.

(Note that  $P_{ij}^{-1} \neq 1/P_{ij}$ .) Also,

$$\begin{aligned}
V[\text{Cov}_C(\hat{I}, H)] &\cong V[b' \Delta p b] + V[b' \Delta g_W] - 2\text{Cov}[b' \Delta p b, b' \Delta g_W] \\
&= \sum_{ijk} b_i b_j b_k b_m \text{Cov}[\hat{P}_{ij}, \hat{P}_{km}] + \sum_{ij} b_i b_j \text{Cov}[\hat{G}_{iW}, \hat{G}_{jW}] \\
&\quad - 2 \sum_{ijk} b_i b_j b_k \text{Cov}[\hat{P}_{ij}, \hat{G}_{kW}].
\end{aligned}$$

Similarly,

$$\begin{aligned} v_c(\hat{I}) &= \hat{b}' \hat{P} \hat{b} = \hat{G}_W' \hat{P}^{-1} \hat{P} \hat{P}^{-1} \hat{G}_W \cong (G_W' + \Delta g_W') (I_n - P^{-1} \Delta p \\ &\quad + P^{-1} \Delta p P^{-1} \Delta p) P^{-1} P (I_n - P^{-1} \Delta p + P^{-1} \Delta p P^{-1} \Delta p) P^{-1} (G_W \\ &\quad + \Delta g_W) \\ &\cong b' G_W - 2b' \Delta p b + 3b' \Delta p P^{-1} \Delta p b + 2b' \Delta g_W \\ &\quad - 4b' \Delta p P^{-1} \Delta g_W + \Delta g_W' P^{-1} \Delta g_W . \end{aligned}$$

$$\begin{aligned} \text{So } E[V_C(\hat{I})] &\cong b' G_W + 3b' E[\Delta p P^{-1} \Delta p] b \\ &\quad - 4b' E[\Delta p P^{-1} \Delta g_W] + E[\Delta g_W' P^{-1} \Delta g_W] \\ &= V(I) + 3 \sum_{ijk} b_i P_{jk}^{-1} b_m \text{Cov}[\hat{p}_{ij}, \hat{p}_{km}] \\ &\quad - 4 \sum_{ijk} b_i P_{jk} \text{Cov}[\hat{p}_{ij}, \hat{g}_{kW}] \\ &\quad + \sum_{ij} P_{ij}^{-1} \text{Cov}[\hat{g}_{iW}, \hat{g}_{jW}] \end{aligned}$$

$$\begin{aligned} \text{and } v[V_C(\hat{I})] &\cong 4V[b' \Delta p b] + 4V[b' \Delta g_W] - 8 \text{Cov}[b' \Delta p b, b' \Delta g_W] \\ &\cong 4 \sum_{ijkm} b_i b_j b_k b_m \text{Cov}[\hat{p}_{ij}, \hat{p}_{km}] \\ &+ 4 \sum_{ij} b_i b_j \text{Cov}[\hat{g}_{iW}, \hat{g}_{jW}] \\ &- 8 \sum_{ijk} b_i b_j b_k \text{Cov}[\hat{p}_{ij}, \hat{g}_{kW}] \end{aligned}$$

The last term needed for evaluation of  $E[\Delta H']$  and  $V[\Delta H']$  is

$$\begin{aligned} \text{Cov}[\text{Cov}_c(\hat{I}, H), v_c(\hat{I})] &= \text{Cov}[G_W' \hat{b}, \hat{b}' P \hat{b}] \\ &\cong \text{Cov}[(b' G_W - b' \Delta p b + b' \Delta g_W), (b' G_W + 2\Delta g_W' b - 2b' \Delta p b)] \\ &= 2V[b' \Delta p b] + 2V[b' \Delta g_W] - 4 \text{Cov}[b' \Delta p b, b' \Delta g_W] \end{aligned}$$

$$= 2 \sum_{ijk} b_i b_j b_k b_m \text{Cov}[\hat{P}_{ij}, \hat{P}_{km}] + 2 \sum_{ij} b_i b_j \text{Cov}[\hat{G}_{iW}, \hat{G}_{jW}]$$

$$- 4 \sum_{ijk} b_i b_j b_k \text{Cov}[\hat{P}_{ij}, \hat{G}_{kW}].$$

Note that  $2 \text{Cov}[\text{Cov}_C(\hat{I}, H), V_C(\hat{I})] \cong V[V_C(\hat{I})] \cong 4V[\text{Cov}_C(\hat{I}, H)]$ .

For the special case when

$$H = \sum_i a_i G_i \text{ and thus } G_{iW} = \sum_j a_j G_{ij} \text{ and } \hat{G}_{iW} = \sum_j a_j \hat{G}_{ij}$$

or, in matrix notation,

$$G_W = Ga \text{ and } \hat{G}_W = \hat{G}a,$$

the above equations are altered as follows:

$$E[\text{Cov}_C(\hat{I}, H)] \cong \text{Cov}(I, H) + \sum_{ijk} b_i P_{jk}^{-1} b_m \text{Cov}[\hat{P}_{ij}, \hat{P}_{km}]$$

$$- \sum_{ijk} b_i P_{jk}^{-1} a_m \text{Cov}[\hat{P}_{ij}, \hat{G}_{km}],$$

$$V[\text{Cov}_C(\hat{I}, H)] \cong \sum_{ijk} b_i b_j b_k b_m \text{Cov}[\hat{P}_{ij}, \hat{P}_{jm}]$$

$$+ \sum_{ijk} b_i a_j b_k a_m \text{Cov}[\hat{G}_{ij}, \hat{G}_{km}]$$

$$- 2 \sum_{ijk} b_i b_j b_k a_m \text{Cov}[\hat{P}_{ij}, \hat{G}_{km}],$$

$$E[V_C(\hat{I})] = V(I) + 3 \sum_{ijk} b_i P_{jk}^{-1} b_m \text{Cov}[\hat{P}_{ij}, \hat{P}_{km}]$$

$$- 4 \sum_{ijk} b_i P_{jk} a_m \text{Cov}[\hat{P}_{ij}, \hat{G}_{km}]$$

$$+ \sum_{ijk} a_i P_{jk}^{-1} a_m \text{Cov}[\hat{G}_{ij}, \hat{G}_{km}]$$

$$\begin{aligned}
V[V_C(\hat{I})] &= 4 \sum_{ijklm} b_i b_j b_k b_m \text{Cov}[\hat{P}_{ij}, \hat{P}_{km}] \\
&+ 4 \sum_{ijklm} b_i a_j b_k a_m \text{Cov}[\hat{G}_{ij}, \hat{G}_{km}] \\
&- 8 \sum_{ijklm} b_i b_j b_k a_m \text{Cov}[\hat{P}_{ij}, \hat{G}_{km}],
\end{aligned}$$

and

$$\begin{aligned}
\text{Cov}[\text{Cov}_C(\hat{I}, H), V_C(\hat{I})] &\approx 2 \sum_{ijklm} b_i b_j b_k b_m \text{Cov}[\hat{P}_{ij}, \hat{P}_{km}] \\
&+ 2 \sum_{ijklm} b_i a_j b_k a_m \text{Cov}[\hat{G}_{ij}, \hat{G}_{km}] \\
&- 4 \sum_{ijklm} b_i b_j b_k a_m \text{Cov}[\hat{P}_{ij}, \hat{G}_{km}].
\end{aligned}$$

The derivations of these approximations were carried out in such a manner that all expectations, variances, and covariances were retained in the equations, while the moments of higher order than these were neglected. It is difficult to assess the reliability of these approximations or the reliability of the approximations for  $E[\Delta H']$  and  $V[\Delta H']$ . The theory underlying the approximation procedure indicates that, when the errors in estimating the  $G_{ij}$  and  $P_{ij}$  values are small, the approximations developed in this section will be close to the true values. Of course, the fact that these errors are not always small is the reason for investigating the entire problem and the reliability of these approximations

is questionable and will be considered further in a later section.

The phrase "effectiveness of an estimation procedure" will be used in this thesis to indicate the extent to which a certain estimation procedure yields estimates which will result in progress near the maximum attainable progress when an index calculated from these estimates is used for selection. The term "estimation procedure" means the method of estimation--including the type of analysis and the quantity of data--used to obtain the estimates.

The second moment of the deviations of expected progress of calculated indexes from the maximum attainable progress may be used as a measure of the effectiveness of a particular estimation procedure. This second moment is

$$E[(\Delta H' - \Delta H)^2] = V[\Delta H'] + [E[\Delta H'] - \Delta H]^2$$

with the smaller values indicating the more effective estimation procedures. An alternative would be to take the first moment,  $\Delta H - E[\Delta H']$ , as the measure of the effectiveness of the estimation procedures. However,  $V[\Delta H'] + [E[\Delta H'] - \Delta H]^2$  seems preferable since this measure puts increased emphasis against the more extreme values of  $E[\Delta H']$  which are likely to be small or negative. However, due to the similarity of the two measures, it seems likely that in most cases the two measures would rank different estimation procedures in much the same order. These two terms might also be called the mean and mean squared decrease from maximum progress. The



estimation procedures which could be compared would be those utilizing different types of analyses of data, such as daughter-dam covariances versus paternal-half-sib analyses of variance and covariance, or those involving different arrangements of the data for a particular type of analysis, such as 50 sires with 20 offspring each versus 100 sires with 10 offspring each.

The formulae presented earlier are completely general for any type of estimation procedures. There remains the problem of evaluating the variances and covariances of the estimates of the various genetic and phenotypic parameters which are used in calculating the index.

The assumption of unbiased estimators was necessary in order to simplify the equations to a form which even then remains barely simple enough to be usable. The increased complexity necessary to allow for biased estimators seems to be almost insurmountable.

Another facet of index selection which may be considered is the accuracy of using the estimates of the parameters in the equation for expected progress. Whereas the progress expected from selection when the true parameters are known is

$$\Delta H = \bar{i}_S B_{HI} \sigma_I = \bar{i}_S (b' G_W)^{\frac{1}{2}},$$

in actual practice estimated values are substituted for parameters, thus obtaining

$$\hat{\Delta H} = \bar{i}_S (\hat{b}' \hat{G}_W)^{\frac{1}{2}}.$$

This value has an expectation

$$E[\hat{\Delta} H] = \bar{i}_s E[\hat{b}' \hat{G}_W]^{\frac{1}{2}} \cong \bar{i}_s E^{\frac{1}{2}}[\hat{b}' \hat{G}_W] \left[ 1 - \frac{v[\hat{b}' \hat{G}_W]}{8E^2[\hat{b}' \hat{G}_W]} \right]$$

and a variance,

$$v[\hat{\Delta} H] = \bar{i}_s^2 v[\hat{b}' \hat{G}_W]^{\frac{1}{2}} \cong \bar{i}_s^2 \frac{v[\hat{b}' \hat{G}_W]}{4E[\hat{b}' \hat{G}_W]}$$

Since  $\hat{b}' \hat{G}_W = \hat{G}_W' \hat{P}^{-1} \hat{G}_W \cong (G_W' + \Delta g_W') (I - P^{-1} \Delta p$

$$\begin{aligned} & + P^{-1} \Delta p P^{-1} \Delta p) P^{-1} (G_W + \Delta g_W) \\ \cong & G_W' P^{-1} G_W - G_W' P^{-1} \Delta p P^{-1} G_W + G_W' P^{-1} \Delta p P^{-1} \Delta p P^{-1} G_W \\ & + G_W' P^{-1} \Delta g_W - G_W' P^{-1} \Delta p P^{-1} \Delta g_W \\ & + \Delta g_W' P^{-1} G_W - \Delta g_W' P^{-1} \Delta p P^{-1} G_W \\ & + \Delta g_W' P^{-1} \Delta g_W, \\ = & \text{Cov}(I, H) - b' \Delta p b + b' \Delta p P^{-1} \Delta p b + 2b' \Delta g_W \\ & - 2b \Delta p P^{-1} \Delta g_W + \Delta g_W' P^{-1} \Delta g_W \end{aligned}$$

$$\begin{aligned} E[\hat{b}' \hat{G}_W] & \cong \text{Cov}(I, H) + b' E[\Delta p P^{-1} \Delta p] b - 2b E[\Delta p P^{-1} \Delta g_W] \\ & + E[\Delta g_W' P^{-1} \Delta g_W] \\ = & \text{Cov}(I, H) + \sum_{ijk} b_i P_{jk}^{-1} b_m \text{Cov}[\hat{P}_{ij}, \hat{P}_{km}] \\ & - 2 \sum_{ijk} b_i P_{jk}^{-1} \text{Cov}[\hat{P}_{ij}, \hat{G}_{kW}] \\ & + \sum_{ij} P_{ij}^{-1} \text{Cov}[\hat{G}_{iW}, \hat{G}_{jW}] \end{aligned}$$

$$\begin{aligned} \text{and } v[\hat{b}' \hat{G}_W] & \cong v[b' \Delta p b] + 4v[b' \Delta g_W] - 4 \text{Cov}[b' \Delta p b, b \Delta g_W] \\ = & \sum_{ijk} b_i b_j b_k b_m \text{Cov}[\hat{P}_{ij}, \hat{P}_{km}] + 4 \sum_{ij} b_i b_j \text{Cov}[\hat{G}_{iW}, \hat{G}_{jW}] \\ & - 4 \sum_{ijk} b_i b_j b_k \text{Cov}[\hat{P}_{ij}, \hat{G}_{kW}]. \end{aligned}$$

For the case when

$$H = \sum_i a_i G_i,$$

these values are altered as follows:

$$\begin{aligned} E[\hat{b}' \hat{G}a] &= \text{Cov}(I, H) + \sum_{ijk m} b_i p_{jk}^{-1} b_m \text{Cov}[\hat{p}_{ij}, \hat{p}_{km}] \\ &\quad - 2 \sum_{ijk m} b_i p_{jk}^{-1} a_m \text{Cov}[\hat{p}_{ij}, \hat{G}_{km}] \\ &\quad + \sum_{ijk m} a_i p_{jk}^{-1} a_m \text{Cov}[\hat{G}_{ij}, \hat{G}_{km}] \end{aligned}$$

and

$$\begin{aligned} V[\hat{b}' \hat{G}a] &= \sum_{ijk m} b_i b_j b_k b_m \text{Cov}[\hat{p}_{ij}, \hat{p}_{km}] \\ &\quad + 4 \sum_{ijk m} b_i a_j b_k a_m \text{Cov}[\hat{G}_{ij}, \hat{G}_{km}] \\ &\quad - 4 \sum_{ijk m} b_i b_j b_k a_m \text{Cov}[\hat{p}_{ij}, \hat{G}_{km}]. \end{aligned}$$

Another quantity of interest is  $\text{Cov}[\hat{\Delta}H, \Delta H']$ . Using the approximations presented earlier, it is found that

$$\begin{aligned} \text{Cov}[\hat{\Delta}H, \Delta H'] &= \bar{I}_S^2 \text{Cov}[(\hat{b}' \hat{G}_W)^{\frac{1}{2}}, \frac{\text{Cov}_C(\hat{I}, H)}{V_C^{\frac{1}{2}}(\hat{I})}] \\ &\cong \bar{I}_S^2 \left\{ \frac{\text{Cov}[\hat{b}' \hat{G}_W, \text{Cov}_C(\hat{I}, H)]}{2E^{\frac{1}{2}}[\hat{b}' \hat{G}_W] E^{\frac{1}{2}}[V_C(\hat{I})] \left[ 1 - \frac{V[V_C(\hat{I})]}{8E^2[V_C(\hat{I})]} \right]} \right. \\ &\quad \left. - \frac{E[\text{Cov}_C(\hat{I}, H)] \text{Cov}[\hat{b}' \hat{G}_W, V_C(\hat{I})]}{4E^{\frac{1}{2}}[\hat{b}' \hat{G}_W] E^{3/2}[V_C(\hat{I})] \left[ 1 - \frac{V[V_C(\hat{I})]}{8E^2[V_C(\hat{I})]} \right]^2} \right\} \end{aligned}$$

This expression involves two quantities not previously presented. These are

$$\begin{aligned}
 \text{Cov}[\hat{b}' \hat{G}_W, \text{Cov}_C(\hat{I}, H)] &= \text{Cov}[\hat{G}_W' \hat{P}^{-1} \hat{G}_W, \hat{G}_W' \hat{P}^{-1} G_W] \\
 &\cong V[b' \Delta p b] - 3 \text{Cov}[b' \Delta p b, b' \Delta g_W] + 2V[b' \Delta g_W] \\
 &= \sum_{ijklm} b_i b_j b_k b_m \text{Cov}[\hat{P}_{ij}, \hat{P}_{km}] \\
 &\quad - 3 \sum_{ijk} b_i b_j b_k \text{Cov}[\hat{P}_{ij}, \hat{G}_{kW}] + 2 \sum_{ij} b_i b_j \text{Cov}[\hat{G}_{iW}, \hat{G}_{jW}]
 \end{aligned}$$

$$\begin{aligned}
 \text{and } \text{Cov}[\hat{b}' \hat{G}_W, V_C(\hat{I})] &= \text{Cov}[\hat{G}_W' \hat{P}^{-1} \hat{G}_W, \hat{G}_W' \hat{P}^{-1} P \hat{P}^{-1} \hat{G}_W] \\
 &\cong 2V[b' \Delta p b] - 6 \text{Cov}[b' \Delta p b, b' \Delta g_W] + 4V[b' \Delta g_W] \\
 &= 2 \sum_{ijklm} b_i b_j b_k b_m \text{Cov}[\hat{P}_{ij}, \hat{P}_{km}] \\
 &\quad - 6 \sum_{ijk} b_i b_j b_k \text{Cov}[\hat{P}_{ij}, \hat{G}_{kW}] + 4 \sum_{ij} b_i b_j \text{Cov}[\hat{G}_{iW}, \hat{G}_{jW}].
 \end{aligned}$$

These expressions are modified for the case when

$$H = \sum_i a_i G_i$$

to become

$$\begin{aligned}
 \text{Cov}[\hat{b}' \hat{G}_a, \text{Cov}_C(\hat{I}, H)] &\cong \sum_{ijklm} b_i b_j b_k b_m \text{Cov}[\hat{P}_{ij}, \hat{P}_{km}] \\
 &\quad - 3 \sum_{ijklm} b_i b_j b_k a_m \text{Cov}[\hat{P}_{ij}, \hat{G}_{km}] \\
 &\quad + 2 \sum_{ijklm} b_i a_j b_k a_m \text{Cov}[\hat{G}_{ij}, \hat{G}_{km}]
 \end{aligned}$$

and

$$\begin{aligned}
\text{Cov}[\hat{b}', \hat{G}_a, v_c(\hat{I})] &\approx 2 \sum_{ijk m} b_i b_j b_k b_m \text{Cov}[\hat{P}_{ij}, \hat{P}_{km}] \\
&- 6 \sum_{ijk m} b_i b_j b_k a_m \text{Cov}[\hat{P}_{ij}, \hat{G}_{km}] \\
&+ 4 \sum_{ijk m} b_i a_j b_k a_m \text{Cov}[\hat{G}_{ij}, \hat{G}_{km}].
\end{aligned}$$

The bias and accuracy of  $\hat{\Delta H}$ , which is an estimator of  $\Delta H'$ , may be evaluated by considering the two values

$$\begin{aligned}
&E[\hat{\Delta H}] - E[\Delta H'] \text{ and } E[(\hat{\Delta H} - \Delta H')^2] = \\
&\left[ E[\hat{\Delta H}] - E[\Delta H'] \right]^2 + V[\hat{\Delta H}] + V[\Delta H'] - 2 \text{Cov}[\hat{\Delta H}, \Delta H'].
\end{aligned}$$

The first of these values, the mean deviation of  $\hat{\Delta H}$  from  $\Delta H'$ , gives the amount by which  $\hat{\Delta H}$  over- or under-estimates  $\Delta H'$  if such a tendency exists, while the latter value is the mean squared difference between  $\hat{\Delta H}$  and  $\Delta H'$ .

The magnitude of the mean squared difference between  $\hat{\Delta H}$  and  $\Delta H'$  is dependent upon the mean difference, the difference in the variances and the correlation between the two variables since

$$\begin{aligned}
E[(\hat{\Delta H} - \Delta H')^2] &= 2 \left[ \frac{V[\hat{\Delta H}] + V[\Delta H']}{2} \right. \\
&\quad \left. - R[\hat{\Delta H}, \Delta H'] V^{\frac{1}{2}}[\hat{\Delta H}] V^{\frac{1}{2}}[\Delta H'] \right] \\
&\quad + \left[ E[\hat{\Delta H}] - E[\Delta H'] \right]^2
\end{aligned}$$

$$\text{where } R[\hat{\Delta H}, \Delta H'] = \frac{\text{Cov}[\hat{\Delta H}, \Delta H']}{V^{\frac{1}{2}}[\hat{\Delta H}] V^{\frac{1}{2}}[\Delta H']}.$$

The effect of the difference in the variances is evident because  $V^{\frac{1}{2}}[\hat{\Delta}H] V^{\frac{1}{2}}[\Delta H']$  becomes increasingly less than  $\frac{V[\hat{\Delta}H] + V[\Delta H']}{2}$  as the difference between  $V[\hat{\Delta}H]$  and  $V[\Delta H']$  increases in absolute value.

A restatement of the definitions of the symbols  $\Delta H$ ,  $\Delta H'$ , and  $\hat{\Delta}H$ , seems desirable for clarification at this point. The symbol  $\Delta H$  represents the maximum attainable progress; that is, the progress attainable if the true population parameters were known and were used for the calculation of an index serving as a basis for selection. The symbol  $\Delta H'$  indicates the expected progress which would result when a particular index,  $\hat{I}$ , based upon estimates is used for selection. Both  $\Delta H$  and  $\Delta H'$  represent the expected values for progress; that is, the progress that would result if selection was carried out in an infinite population. The variation in progress due to the finiteness of the population in which selection is carried out has been excluded from consideration in this study. The other symbol,  $\hat{\Delta}H$ , represents the estimate of progress which would be made based upon the estimates of the population parameters when the true parameters, which determine  $\Delta H$  and  $\Delta H'$ , are not known. The two main points of consideration in this thesis are the closeness of the population of  $\Delta H'$  values to the  $\Delta H$  value for each set of true parameter values and each estimation procedure and the closeness of  $\hat{\Delta}H$  values to the corresponding  $\Delta H'$  values.

## VARIANCES AND COVARIANCES OF ESTIMATES

Examination of the equations developed in the preceding section shows, as would be expected, that the decrease,  $\Delta\hat{H} - \Delta H_{1/2}$ , of expected progress for calculated indexes from the maximum attainable progress is dependent upon the magnitude of the sampling variances and covariances of the estimates. Also, the accuracy of estimating expected progress,  $\hat{\Delta H}$ , is dependent upon these values. It is therefore necessary to evaluate these variances and covariances in order to compare the effectiveness of different estimation procedures.

The standard errors of estimates of genetic correlations were investigated by Rae (1950), Reeve (1955), Robertson (1959a), and Tallis (1959), each author considering a specific procedure of estimating genetic correlations. The standard errors of heritability estimates were studied by Osborne and Patterson (1952) and Robertson (1959b) for an analysis of variance procedure involving sires, dams within sires, and within families classifications, by Tallis and Klosterman (1959) for an analysis of variance procedure involving paternal-half-sibs, and by Van Vleck, Searle, and Henderson (1960) and Latter and Robertson (1960) for regression methods.

The present paper is restricted to studying estimations from the paternal-half-sib variance and covariance procedure.

Since primary interest is directed to the estimates of the additive genetic and phenotypic variances and covariances, the above mentioned studies on estimates of heritabilities and genetic correlations are not directly applicable here. However, these studies were enlightening and stimulating to the present problem.

Kendall and Stuart (1958) reviewed the theory that leads to the general result

$$\text{Cov}[m_{ij}, m_{pq}] = \frac{\mu_{ip} \mu_{jq} + \mu_{iq} \mu_{jp}}{n - 1}$$

where  $m_{ij}$  is a sample covariance between the variables  $x_i$  and  $x_j$  based on  $n$  observations

and  $\mu_{ij}$  is the corresponding population covariance.

For the derivation of this equation, it is assumed that the variables,  $x_i$ ,  $x_j$ ,  $x_p$ , and  $x_q$ , are distributed in a multivariate normal distribution. This equation is general and includes the special cases

$$V[m_{ii}] = \frac{2}{n-1} \mu_{ii}^2 \text{ when } i = j = p = q,$$

$$V[m_{ij}] = \frac{1}{n-1} [\mu_{ii} \mu_{jj} + \mu_{ij}^2] \text{ when } i = p \text{ and } j = q,$$

$$\text{Cov}[m_{ii}, m_{pp}] = \frac{2}{n-1} \mu_{ip}^2 \text{ when } i = j \text{ and } p = q,$$

etc.

This leads to the equations for the variances and covariances of the mean squares and mean products of the sire progeny



group analyses for traits  $i$ ,  $j$ ,  $p$ , and  $q$  for the special case of equal number ( $m$ ) of offspring in each sire group as follows:

$$\text{Cov}[w_{ij}, w_{pq}] = \frac{1}{s(m-1)} [\sigma_{ipe} \sigma_{jqe} + \sigma_{iqe} \sigma_{jpe}]$$

$$\begin{aligned} \text{and } \text{Cov}[s_{ij}, s_{pq}] &= \frac{1}{s-1} [(\sigma_{ipe} + m\sigma_{ips})(\sigma_{jqe} + m\sigma_{jqs}) \\ &\quad + (\sigma_{iqe} + m\sigma_{iqs})(\sigma_{jpe} + m\sigma_{jps})] \end{aligned}$$

with the resulting special cases when some of the traits denoted by  $i$ ,  $j$ ,  $p$ , and  $q$  are the same. The assumption that the sire elements and the random elements for each trait are normally distributed is necessary. Also

$$\text{Cov}[w_{ij}, s_{pq}] = 0.$$

The situation involving unequal numbers of offspring for the different sire groups is complicated considerably. Crump (1947) developed the equations for the variances of mean squares for groups and within groups, which in the symbolism used in the present paper are

$$V(w_{ii}) = \frac{2\sigma_{iie}^2}{M-s}$$

and

$$\begin{aligned} V(s_{ii}) &= 2 \left[ \left[ \frac{1}{m} \sum_k m_k^2 \left( \sigma_{iis} + \frac{\sigma_{iie}}{m_k} \right) \right]^2 \right. \\ &\quad \left. + \sum_k \left[ m_k \left( \sigma_{iis} + \frac{\sigma_{iie}}{m_k} \right) \right]^2 - \frac{2}{M} \sum_k m_k^3 \left( \sigma_{iis} + \frac{\sigma_{iie}}{m_k} \right)^2 \right]. \end{aligned}$$

Again, the assumption of normality is necessary in the derivation.

For the purposes of this problem, a more general equation that will give the variances and covariances of the sire mean squares and mean products for four traits is needed. This can be accomplished by some rather tedious algebra starting with models for the phenotypic values of the four traits as follows:

$$x_{ikh} = a_k + e_{kh},$$

$$x_{jkh} = \beta_1 a_k + b_k + \beta_2 e_{kh} + f_{kh},$$

$$x_{pkh} = \beta_3 a_k + \beta_4 b_k + c_k + \beta_5 e_{kh} + \beta_6 f_{kh} + g_{kh},$$

$$\text{and } x_{qkh} = \beta_7 a_k + \beta_8 b_k + \beta_9 c_k + d_k + \beta_{10} e_{kh} + \beta_{11} f_{kh} \\ + \beta_{12} g_{kh} + h_{kh},$$

where  $a_k$ ,  $b_k$ ,  $c_k$ ,  $d_k$ ,  $e_{kh}$ ,  $f_{kh}$ ,  $g_{kh}$ , and  $h_{kh}$  are independent variables which are normally distributed. The subscript  $k$  denotes the different sire groups in the experiment and takes on values 1 to  $s$ , the number of sire groups. The subscript  $h$  denotes a particular offspring in a particular sire group and takes on the values from 1 to  $m_k$ , the number of offspring in the  $k^{\text{th}}$  sire group. The  $\beta$  values are constants which, along with the variances of the independent variables, determine the parameters of the multivariate distribution of the sire elements and random elements for the four traits.

The parameters of this multivariate distribution are

$$\sigma_{ijs} = \beta_1 V(a_k), \quad \sigma_{ije} = \beta_2 V(e_{kh}),$$

$$\sigma_{ips} = \beta_3 v(a_k), \sigma_{ipe} = \beta_5 v(e_{kh}),$$

$$\sigma_{jps} = \beta_1 \beta_3 v(a_k) + \beta_4 v(b_k),$$

$$\sigma_{jpe} = \beta_2 \beta_5 v(e_{kh}) + \beta_6 v(f_{kh}),$$

$$\sigma_{jqs} = \beta_1 \beta_7 v(a_k) + \beta_8 v(b_k),$$

$$\sigma_{jqe} = \beta_2 \beta_{10} v(e_{kh}) + \beta_{11} v(f_{kh}),$$

$$\sigma_{pqs} = \beta_3 \beta_7 v(a_k) + \beta_4 \beta_8 v(b_k) + \beta_9 v(c_k),$$

$$\text{and } \sigma_{pqe} = \beta_5 \beta_{10} v(e_{kh}) + \beta_6 \beta_{11} v(f_{kh}) + \beta_{12} v(g_{kh}).$$

$$\text{If } s_{ij} = \frac{1}{s-1} \left[ \sum_k \frac{\left( \sum_h x_{ikh} \right) \left( \sum_h x_{jkh} \right)}{m_k} - \frac{\left( \sum_{kh} x_{ikh} \right) \left( \sum_{kh} x_{jkh} \right)}{M} \right]$$

$$\text{and } s_{pq} = \frac{1}{s-1} \left[ \sum_k \frac{\left( \sum_h x_{pkh} \right) \left( \sum_h x_{qkh} \right)}{m_k} - \frac{\left( \sum_{kh} x_{pkh} \right) \left( \sum_{kh} x_{qkh} \right)}{M} \right]$$

are expanded in terms of the  $\beta$  values and the independent normally distributed variables and  $\text{Cov}[s_{ij}, s_{pq}]$  is evaluated, it is found that

$$\begin{aligned} \text{Cov}[s_{ij}, s_{pq}] &= \frac{K}{(s-1)^2} [2 \beta_1 \beta_3 \beta_7 v^2(a_k) \\ &+ \beta_3 \beta_8 v(a_k) v(b_k) + \beta_4 \beta_7 v(a_k) v(b_k)] \\ &+ \frac{m_0}{s-1} [\beta_1 \beta_3 \beta_{10} v(a_k) v(e_{kh}) + \beta_2 \beta_3 \beta_{10} v(a_k) v(e_{kh}) \\ &+ \beta_1 \beta_5 \beta_7 v(a_k) v(e_{kh}) + \beta_2 \beta_5 \beta_7 v(a_k) v(e_{kh}) \\ &+ \beta_3 \beta_{11} v(a_k) v(f_{kh}) + \beta_6 \beta_7 v(a_k) v(f_{kh}) \\ &+ \beta_4 \beta_{10} v(b_k) v(e_{kh}) + \beta_5 \beta_8 v(b_k) v(e_{kh})] \\ &+ \frac{1}{s-1} [2 \beta_2 \beta_5 \beta_{10} v^2(e_{kh}) + \beta_5 \beta_{11} v(e_{kh}) v(f_{kh}) \\ &+ \beta_6 \beta_{10} v(e_{kh}) v(f_{kh})] \end{aligned}$$

where

$$K = \sum_k m_k^2 - 2 \frac{\sum_k m_k^3}{M} + \frac{(\sum_k m_k^2)^2}{M^2}$$

The value K must always be positive since it is derived to be equal to

$$\frac{V \left[ \sum_k m_k a_k^2 - \frac{(\sum_k m_k a_k)^2}{M} \right]}{2V^2[a_k]}$$

Substituting the population parameters into the above equation, it simplifies to

$$\begin{aligned} \text{Cov}[s_{ij}, s_{pq}] = & \frac{K}{(s-1)^2} [\sigma_{ips} \sigma_{jqe} + \sigma_{iqe} \sigma_{jps}] \\ & + \frac{m_0}{s-1} [\sigma_{ips} \sigma_{jqe} + \sigma_{jqe} \sigma_{ipe} + \sigma_{iqe} \sigma_{jpe} \\ & \quad + \sigma_{jps} \sigma_{iqe}] \\ & + \frac{1}{s-1} [\sigma_{ipe} \sigma_{jqe} + \sigma_{iqe} \sigma_{jps}]. \end{aligned}$$

This equation is also a general equation, and all the variances and covariances among sire mean squares and mean products for any number of traits may be derived from it. The equation is consistent with the equation presented earlier for  $\text{Cov}[s_{ij}, s_{pq}]$  when there are equal numbers of offspring in the sire groups. It is also consistent with Crump's (1947) equation for the variance of a group mean square.

The general equation presented earlier for the covariance of sample covariances shows that

$$\text{Cov}[w_{ij}(k), w_{pq}(k)] = \frac{\sigma_{ipe} \sigma_{jqe} + \sigma_{iqe} \sigma_{jpe}}{m_k - 1}$$

where  $w_{ij}(k)$  is the sample covariance between traits  $i$  and  $j$  in the  $k^{\text{th}}$  sire group.

Since

$$w_{ij} = \frac{\sum_k (m_k - 1) w_{ij}(k)}{\sum_k (m_k - 1)},$$

$$\text{Cov}[w_{ij}, w_{pq}] = \text{Cov} \left[ \frac{\sum_k (m_k - 1) w_{ij}(k)}{\sum_k (m_k - 1)}, \frac{\sum_k (m_k - 1) w_{pq}(k)}{\sum_k (m_k - 1)} \right]$$

$$= \frac{1}{(M-s)^2} \sum_k (m_k - 1)^2 \text{Cov}[w_{ij}(k), w_{pq}(k)]$$

$$= \frac{1}{(M-s)^2} \sum_k (m_k - 1) [\sigma_{ipe} \sigma_{jqe} + \sigma_{iqe} \sigma_{jpe}]$$

$$= \frac{1}{M-s} [\sigma_{ipe} \sigma_{jqe} + \sigma_{iqe} \sigma_{jpe}]$$

where homogeneity of variation or covariation within groups is assumed and

$$\text{Cov}[w_{ij}(k), w_{pq}(k')] = 0 \text{ for } k \neq k'.$$

Since the arrangement is hierarchical,

$$\text{Cov}[s_{ij}, w_{pq}] = 0$$

in all cases.

Since

$$\begin{aligned}
 t_{ij} &= \frac{s-1}{M-1} s_{ij} + \frac{M-s}{M-1} w_{ij}, \\
 \text{Cov}[t_{ij}, t_{pq}] &= \frac{(s-1)^2}{(M-1)^2} \text{Cov}[s_{ij}, s_{pq}] + \frac{(M-s)^2}{(M-1)^2} \text{Cov}[w_{ij}, w_{pq}] \\
 &= \frac{K}{(M-1)^2} [\sigma_{ips} \sigma_{jqs} + \sigma_{iqs} \sigma_{jps}] \\
 &\quad + \frac{m_0(s-1)}{(M-1)^2} [\sigma_{ips} \sigma_{jqe} + \sigma_{jqs} \sigma_{ipe} + \sigma_{iqs} \sigma_{jpe} \\
 &\quad \quad \quad + \sigma_{jps} \sigma_{iqe}] \\
 &\quad + \frac{1}{M-1} [\sigma_{ipe} \sigma_{jqe} + \sigma_{iqe} \sigma_{jpe}],
 \end{aligned}$$

and since

$$\begin{aligned}
 \hat{\sigma}_{ije} + \hat{\sigma}_{ijs} &= \frac{1}{m_0} s_{ij} + \frac{m_0-1}{m_0} w_{ij}, \\
 \text{Cov}[(\hat{\sigma}_{ije} + \hat{\sigma}_{ijs}), (\hat{\sigma}_{pqe} + \hat{\sigma}_{pqs})] &= \frac{1}{m_0^2} \text{Cov}[s_{ij}, s_{pq}] \\
 &\quad + \frac{(m_0-1)^2}{m_0^2} \text{Cov}[w_{ij}, w_{pq}] = \frac{K}{m_0^2 (s-1)^2} [\sigma_{ips} \sigma_{jqs} + \sigma_{iqs} \sigma_{jps}] \\
 &\quad + \frac{1}{m_0 (s-1)} [\sigma_{ips} \sigma_{jqe} + \sigma_{jqs} \sigma_{ipe} + \sigma_{iqs} \sigma_{jpe} \\
 &\quad \quad \quad + \sigma_{jps} \sigma_{iqe}] \\
 &\quad + \frac{M-s + (m_0-1)^2 (s-1)}{m_0^2 (s-1) (M-s)} [\sigma_{ipe} \sigma_{jqe} + \sigma_{iqe} \sigma_{jpe}].
 \end{aligned}$$

These equations allow the evaluation of the relative accuracy of  $t_{ij}$  and  $\hat{\sigma}_{ije} + \hat{\sigma}_{ijs}$  as estimators of  $P_{ij}$ , the

phenotypic variance ( $i = j$ ) or covariance ( $i \neq j$ ). Using the second moment of the estimator about the true value as a measure of the accuracy of that estimator,

$$\begin{aligned}
 E[\hat{\sigma}_{ije} + \hat{\sigma}_{ijs} - p_{ij}]^2 &= V(\hat{\sigma}_{ije} + \hat{\sigma}_{ijs}) \\
 &= \frac{K}{m_o^2 (s-1)^2} [\sigma_{iis} \sigma_{jjs} + \sigma_{ijs}^2] \\
 &+ \frac{1}{m_o (s-1)} [\sigma_{iis} \sigma_{jje} + \sigma_{iie} \sigma_{jjs} + 2\sigma_{ije} \sigma_{ijs}] \\
 &+ \frac{M-s + (m_o-1)^2 (s-1)}{m_o^2 (s-1) (M-s)} [\sigma_{iie} \sigma_{jje} + \sigma_{ije}^2]
 \end{aligned}$$

with the obvious simplification when  $i = j$ . The second moment of  $t_{ij}$ ,

$$\begin{aligned}
 E[t_{ij} - p_{ij}]^2 &= V(t_{ij}) + [E(t_{ij}) - p_{ij}]^2 \\
 &= \frac{K}{(M-1)^2} \sigma_{iis} \sigma_{jjs} + \frac{m_o (s-1)}{(M-1)^2} [\sigma_{iis} \sigma_{jje} \\
 &\quad + \sigma_{iie} \sigma_{jjs} + 2\sigma_{ije} \sigma_{ijs}] \\
 &\quad + \frac{1}{M-1} [\sigma_{iie} \sigma_{jje} + \sigma_{ije}^2] + \left[ \frac{K}{(M-1)^2} + \frac{(M-\sum m_k^2)^2}{M^2 (M-1)^2} \right] \sigma_{ijs}^2.
 \end{aligned}$$

Comparison of the second moments about the true value of these estimators shows that when there are equal numbers of offspring in each sire group,  $t_{ij}$  is a more accurate estimator of  $p_{ij}$  than  $\hat{\sigma}_{ije} + \hat{\sigma}_{ijs}$  since the second moment of  $t_{ij}$  about

$P_{ij}$  is smaller than this value for  $\hat{\sigma}_{ije} + \hat{\sigma}_{ijs}$ , regardless of the parameters of the population. However, for the case of unequal subclass numbers, it is difficult to establish a general conclusion about the relative accuracy of these two estimators. The relative sizes of the second moments of the two estimators are dependent upon  $s$ ,  $M$ ,  $\sum_k m_k^2$ ,  $\sum_k m_k^3$ , and the population parameters, and there does not seem to be any simple solution to indicate the regions where each estimator is more accurate. However, the fact that  $t_{ij}$  is the most reliable for equal subclass numbers indicates that this estimator will also be more accurate when  $V(m_i)$  is small.

Since  $\sigma_{ijs} = \frac{1}{4}G_{ij}$  and  $\sigma_{ije} = P_{ij} - \frac{1}{4}G_{ij}$ , the general formula for the covariances of the mean products may also be expressed in terms of the genetic and phenotypic parameters as follows:

$$\begin{aligned} \text{Cov}[s_{ij}, s_{pq}] &= \frac{K - (2m_0 - 1)(s - 1)}{16(s - 1)^2} [G_{ip} G_{jq} + G_{iq} G_{jp}] \\ &+ \frac{m_0 - 1}{4(s - 1)} [P_{ip} G_{jq} + P_{jq} G_{ip} + P_{jp} G_{iq} + P_{iq} G_{jp}] \\ &+ \frac{1}{s - 1} [P_{ip} P_{jq} + P_{iq} P_{jp}] \end{aligned}$$

$$\begin{aligned} \text{and } \text{Cov}[w_{ij}, w_{pq}] &= \frac{1}{16(M - s)} [G_{ip} G_{jq} + G_{iq} G_{jp}] \\ &- \frac{1}{4(M - s)} [P_{ip} G_{jq} + P_{jq} G_{ip} + P_{iq} G_{jp} + P_{jp} G_{iq}] \\ &+ \frac{1}{M - s} [P_{ip} P_{jq} + P_{iq} P_{jp}]. \end{aligned}$$



These formulae allow the development of general formulae for  $\text{Cov}[\hat{P}_{ij}, \hat{P}_{pq}]$ ,  $\text{Cov}[\hat{G}_{ij}, \hat{P}_{pq}]$ , and  $\text{Cov}[\hat{G}_{ij}, \hat{G}_{pq}]$  in terms of the genetic and phenotypic parameters. When  $G_{ij}$  is estimated by  $\hat{G}_{ij} = 4\hat{a}_{ijs}$  and  $P_{ij}$  is estimated by  $\hat{P}_{ij} = t_{ij}$ ,

$$\begin{aligned} \text{Cov}[\hat{P}_{ij}, \hat{P}_{pq}]_t &= \text{Cov}\left[\left(\frac{s-1}{M-1} s_{ij} + \frac{M-s}{M-1} w_{ij}\right), \left(\frac{s-1}{M-1} s_{pq} + \frac{M-s}{M-1} w_{pq}\right)\right] \\ &= \frac{(s-1)^2}{(M-1)^2} \text{Cov}[s_{ij}, s_{pq}] + \frac{(M-s)^2}{(M-1)^2} \text{Cov}[w_{ij}, w_{pq}] \\ &= \frac{K - 2m_0 (s-1) + M-1}{16(M-1)^2} [G_{ip} G_{jq} + G_{iq} G_{jp}] \\ &\quad + \frac{M - \sum_k m_k^2}{4M(M-1)^2} [P_{ip} G_{jq} + P_{jq} G_{ip} + P_{jp} G_{iq} + P_{iq} G_{jp}] \\ &\quad + \frac{1}{M-1} [P_{ip} P_{jq} + P_{iq} P_{jp}], \end{aligned}$$

$$\begin{aligned} \text{Cov}[G_{ij}, P_{pq}]_t &= \text{Cov}\left[\left(\frac{4}{m_0} s_{ij} - \frac{4}{m_0} w_{ij}\right), \left(\frac{s-1}{M-1} s_{pq} + \frac{M-s}{M-1} w_{pq}\right)\right] \\ &= \frac{4(s-1)}{m_0(M-1)} \text{Cov}[s_{ij}, s_{pq}] - \frac{4(M-s)}{m_0(M-1)} \text{Cov}[w_{ij}, w_{pq}] \\ &= \frac{K - 2m_0 (s-1)}{4m_0 (s-1)(M-1)} [G_{ip} G_{jq} + G_{iq} G_{jp}] \\ &\quad + \frac{1}{M-1} [P_{ip} G_{jq} + P_{jq} G_{ip} + P_{iq} G_{jp} + P_{ip} G_{iq}], \end{aligned}$$

and

$$\begin{aligned}
\text{Cov}[\hat{G}_{ij}, \hat{G}_{pq}] &= \text{Cov} \left[ \left( \frac{4}{m_0} s_{ij} - \frac{4}{m_0} w_{ij} \right), \left( \frac{4}{m_0} s_{pq} - \frac{4}{m_0} w_{pq} \right) \right] \\
&= \frac{16}{m_0^2} \text{Cov}[s_{ij}, s_{pq}] + \frac{16}{m_0^2} \text{Cov}[w_{ij}, w_{pq}] \\
&= \left[ \frac{K - (2m_0 - 1)(s - 1)}{m_0^2 (s - 1)^2} \quad \frac{1}{m_0^2 (M - s)} \right] [G_{ip} G_{jq} + G_{iq} G_{jp}] \\
&\quad + \left[ \frac{4(m_0 - 1)}{m_0^2 (s - 1)} \quad \frac{4}{m_0^2 (M - s)} \right] [P_{ip} G_{jq} + P_{jq} G_{ip} + P_{iq} G_{jp} \\
&\quad \quad \quad + P_{jp} G_{iq}] \\
&\quad + \frac{16(M - 1)}{m_0^2 (M - s)(s - 1)} [P_{ip} P_{jq} + P_{iq} P_{jp}].
\end{aligned}$$

Note the subscript  $t$  to denote that  $\text{Cov}[\hat{P}_{ij}, \hat{P}_{pq}]_t$  and  $\text{Cov}[\hat{G}_{ij}, \hat{P}_{pq}]_t$  represent the covariances when  $P_{ij}$  is estimated by  $\hat{P}_{ij} = t_{ij}$ . When  $P_{ij}$  is estimated by  $\hat{P}_{ij} = \hat{\sigma}_{ije} + \hat{\sigma}_{ijs}$ , the subscript  $\sigma$  will be used on these values.

Since  $t_{ij}$  is a biased estimator of  $P_{ij}$ , the use of these values for the covariances of the estimates in the equations developed in the preceding section is not quite valid. However, when the number of sires is large, the amount of bias is small. For example, with 50 sires with 20 offspring each, the bias is  $.005 G_{ij}$ . Therefore, for such cases the errors introduced by the application of the above theory for unbiased estimators would likely be small due to the smallness of the

bias.

For some circumstances it may be desirable to estimate  $P_{ij}$  by

$$\hat{P}_{ij} = \hat{\sigma}_{ije} + \hat{\sigma}_{ijs}.$$

Then

$$\begin{aligned} \text{Cov}[\hat{P}_{ij}, \hat{P}_{pq}]_{\sigma} &= \text{Cov} \left[ \left( \frac{1}{m_0} s_{ij} + \frac{m_0-1}{m_0} w_{ij} \right), \left( \frac{1}{m_0} s_{pq} + \frac{m_0-1}{m_0} w_{pq} \right) \right] \\ &= \frac{1}{m_0^2} \text{Cov}[s_{ij}, s_{pq}] + \frac{(m_0-1)^2}{m_0^2} \text{Cov}[w_{ij}, w_{pq}] \\ &= \left[ \frac{K - (2m_0-1)(s-1)}{16 m_0^2 (s-1)^2} + \frac{(m_0-1)^2}{16 m_0^2 (M-s)} \right] [G_{ip} G_{jq} + G_{iq} G_{jp}] \\ &\quad + \frac{(m_0-1)(\sum_k m_k^2 - M)}{4 m_0^2 M(M-s)(s-1)} [P_{ip} G_{jq} + P_{jq} G_{ip} + P_{iq} G_{jp} + P_{jp} G_{iq}] \\ &\quad + \left[ \frac{1}{m_0^2 (s-1)} + \frac{(m_0-1)^2}{m_0^2 (M-s)} \right] [P_{ip} P_{jq} + P_{iq} P_{jp}], \end{aligned}$$

and

$$\begin{aligned} \text{Cov}[\hat{G}_{ij}, \hat{P}_{pq}]_{\sigma} &= \text{Cov} \left[ \left( \frac{4}{m_0} s_{ij} - \frac{4}{m_0} w_{ij} \right), \left( \frac{1}{m_0} s_{pq} + \frac{m_0-1}{m_0} w_{pq} \right) \right] \\ &= \frac{4}{m_0^2} \text{Cov}[s_{ij}, s_{pq}] - \frac{4(m_0-1)}{m_0^2} \text{Cov}[w_{ij}, w_{pq}] \\ &= \left[ \frac{K - (2m_0-1)(s-1)}{4 m_0^2 (s-1)^2} - \frac{(m_0-1)}{4 m_0^2 (M-s)} \right] [G_{ip} G_{jq} + G_{iq} G_{jp}] \\ &\quad + \frac{(m_0-1)(M-1)}{m_0^2 (s-1)(M-s)} [P_{ip} G_{jq} + P_{jq} G_{ip} + P_{jp} G_{iq} + P_{iq} G_{jp}] \\ &\quad + \frac{4(\sum_k m_k^2 - M)}{m_0^2 M(s-1)(M-s)} [P_{ip} P_{jq} + P_{iq} P_{jp}]. \end{aligned}$$

SIMULATION OF ESTIMATION FROM MULTI-SIRE SAMPLES  
WITH TWO TRAITS

The equations developed in the preceding sections are based upon large sample theory. Due to their complexity, it is difficult to assess the reliability of these equations for situations involving limited amounts of data. Therefore, an empirical study of the influence of estimation upon index selection was conducted. The purposes of this empirical study were two-fold: (1) to provide an alternate method of determining the influence of sampling errors and (2) to provide a test of the reliability of the approximate equations for situations which might be encountered in practice.

The empirical study was accomplished by simulating samples of paternal-half-sib groups from populations with known genetic and phenotypic parameters on an IBM 650 Data Processing System. A statistical sampling technique such as this is usually termed a "Monte Carlo" technique, and this terminology will be used here.

Wadell and O'Bleness (1960) discussed the simulation of genetic populations on automatic computers. The procedures outlined in their paper were in general followed for the present study, although some minor differences existed. It should be pointed out that the simulation was not a direct simulation of genetic systems as was carried out by Fraser

(1957a and 1957b) and Barker (1958a and 1958b), but was a simulation of the statistical models usually used to describe the genetic systems.

Two phenotypic traits of the  $h^{\text{th}}$  offspring of the  $k^{\text{th}}$  sire may be represented by

$$x_{1kh} = \lambda_1 c_k + \lambda_2 e_{kh}$$

and

$$x_{2kh} = \lambda_2 c_k + \lambda_4 s_k + \lambda_5 e_{kh} + \lambda_6 f_{kh}$$

where  $c_k$ ,  $s_k$ ,  $e_{kh}$ , and  $f_{kh}$  are random independent variables and the  $\lambda$  values are constants. For this problem normally and independently distributed variables with mean zero and unit variance were used for the  $c_k$ ,  $s_k$ ,  $e_{kh}$ , and  $f_{kh}$  values by first obtaining 10 digit uniformly distributed random variables by the power residue method as outlined in IBM Reference Manual; Random Number Generation and Testing (1959) and then converting these to normal variables (except for rounding) by a table look-up procedure using a table for the cumulative normal distribution with a mean of zero and a variance of unity.

The additive genetic variances and covariances and the phenotypic variances and covariances from traits simulated in this manner are functions of the  $\lambda$ -values. These parameters are

$$G_{11} = 4\lambda_1^2, G_{12} = 4\lambda_1 \lambda_3, G_{22} = 4\lambda_3^2 + 4\lambda_4^2,$$

$$P_{11} = \lambda_1^2 + \lambda_2^2, P_{12} = \lambda_1 \lambda_3 + \lambda_2 \lambda_5,$$

$$\text{and } P_{22} = \lambda_3^2 + \lambda_4^2 + \lambda_5^2 + \lambda_6^2.$$

Thus various combinations of additive genetic and phenotypic parameters were simulated by choosing the set of  $\lambda$  values to use with the random variables.

Simulating the usual estimation procedure from analyses of variance and covariance of phenotypes for two traits observed in  $m$  offspring from each of  $s$  sires, the estimates of the additive genetic parameters were calculated as

$$\begin{aligned}\hat{G}_{11} &= \frac{4}{m} \left[ \frac{1}{s-1} \left[ \lambda_1^2 \left[ K_1 - \frac{K_5^2}{sm} \right] + 2\lambda_1 \lambda_2 \left[ K_3 - \frac{K_5 K_6}{sm} \right] \right. \right. \\ &\quad \left. \left. + \lambda_2^2 \left[ \frac{K_4}{m} - \frac{K_6^2}{sm} \right] \right] - \frac{1}{s(m-1)} \left[ \lambda_2^2 \left[ K_2 - \frac{K_4}{m} \right] \right] \right], \\ \hat{G}_{12} &= \frac{4}{m} \left[ \frac{1}{s-1} \left[ \lambda_1 \lambda_3 \left[ K_1 - \frac{K_5^2}{sm} \right] + \lambda_1 \lambda_4 \left[ K_9 - \frac{K_5 K_{15}}{sm} \right] \right. \right. \\ &\quad + (\lambda_1 \lambda_5 + \lambda_2 \lambda_3) \left[ K_3 - \frac{K_5 K_6}{sm} \right] + \lambda_1 \lambda_6 \left[ K_{10} - \frac{K_5 K_{16}}{sm} \right] \\ &\quad + \lambda_2 \lambda_4 \left[ K_{11} - \frac{K_6 K_{15}}{sm} \right] + \lambda_2 \lambda_5 \left[ \frac{K_4}{m} - \frac{K_6^2}{sm} \right] \\ &\quad \left. \left. + \lambda_2 \lambda_6 \left[ \frac{K_{14}}{m} - \frac{K_6 K_{16}}{sm} \right] \right] \right. \\ &\quad \left. - \frac{1}{s(m-1)} \left[ \lambda_2 \lambda_5 \left[ K_2 - \frac{K_4}{m} \right] + \lambda_2 \lambda_6 \left[ K_{13} - \frac{K_{14}}{m} \right] \right] \right],\end{aligned}$$

and

$$\begin{aligned}\hat{G}_{22} &= \frac{4}{m} \left[ \frac{1}{s-1} \left[ \lambda_3^2 \left[ K_1 - \frac{K_5^2}{sm} \right] + \lambda_4^2 \left[ K_7 - \frac{K_{15}^2}{sm} \right] \right. \right. \\ &\quad \left. \left. + \lambda_5^2 \left[ \frac{K_4}{m} - \frac{K_6^2}{sm} \right] + \lambda_6^2 \left[ \frac{K_{17}}{m} - \frac{K_{16}^2}{sm} \right] \right] \right]\end{aligned}$$

$$\begin{aligned}
& + 2\lambda_3 \lambda_4 \left[ K_9 - \frac{K_5 K_{15}}{sm} \right] + 2\lambda_3 \lambda_5 \left[ K_3 - \frac{K_5 K_6}{sm} \right] \\
& + 2\lambda_3 \lambda_6 \left[ K_{10} - \frac{K_5 K_{16}}{sm} \right] + 2\lambda_4 \lambda_5 \left[ K_{11} - \frac{K_{15} K_6}{sm} \right] \\
& + 2\lambda_4 \lambda_6 \left[ K_{12} - \frac{K_{15} K_{16}}{sm} \right] + 2\lambda_5 \lambda_6 \left[ \frac{K_{14}}{m} - \frac{K_6 K_{16}}{sm} \right] \\
& - \frac{1}{s(m-1)} \left[ \lambda_5^2 \left[ K_2 - \frac{K_4}{m} \right] + \lambda_6^2 \left[ K_8 - \frac{K_{17}}{m} \right] \right. \\
& \left. + 2\lambda_5 \lambda_6 \left[ K_{13} - \frac{K_{14}}{m} \right] \right] ,
\end{aligned}$$

where

$$\begin{aligned}
K_1 &= m \sum_k c_k^2, \quad K_2 = \sum_{kh} e_{kh}^2, \\
K_3 &= \sum_k c_k e_k. \quad \text{where } e_k = \sum_h e_{kh}, \\
K_4 &= \sum_k e_k^2, \quad K_5 = m \sum_k c_k, \\
K_6 &= \sum_{kh} e_{kh} = \sum_k e_k, \quad K_7 = m \sum_k s_k^2, \\
K_8 &= \sum_{kh} f_{kh}^2, \quad K_9 = m \sum_k c_k s_k, \\
K_{10} &= \sum_k c_k f_k. \quad \text{where } f_k = \sum_h f_{kh}, \\
K_{11} &= \sum_k s_k e_k, \quad K_{12} = \sum_k s_k f_k, \\
K_{13} &= \sum_{kh} e_{kh} f_{kh}, \quad K_{14} = \sum_k e_k f_k, \\
K_{15} &= m \sum_k s_k, \quad K_{16} = \sum_{kh} f_{kh} = \sum_k f_k, \\
\text{and } K_{17} &= \sum_k f_k^2.
\end{aligned}$$

Similarly, the estimates of the phenotypic variances and covariances as calculated from the total mean squares or products ( $t_{ij}$ 's) were

$$P_{11} = \frac{1}{sm-1} \left[ \lambda_1^2 \left[ K_1 - \frac{K_5^2}{sm} \right] + 2\lambda_1 \lambda_2 \left[ K_3 - \frac{K_5 K_6}{sm} \right] + \lambda_2^2 \left[ K_2 - \frac{K_6^2}{sm} \right] \right]$$

$$P_{12} = \frac{1}{sm-1} \left[ \lambda_1 \lambda_3 \left[ K_1 - \frac{K_5^2}{sm} \right] + \lambda_1 \lambda_4 \left[ K_9 - \frac{K_5 K_{15}}{sm} \right] + \left[ \lambda_1 \lambda_5 + \lambda_2 \lambda_3 \right] \left[ K_3 - \frac{K_5 K_6}{sm} \right] + \lambda_1 \lambda_6 \left[ K_{10} - \frac{K_5 K_{16}}{sm} \right] + \lambda_2 \lambda_4 \left[ K_{11} - \frac{K_6 K_{15}}{sm} \right] + \lambda_2 \lambda_5 \left[ K_2 - \frac{K_6^2}{sm} \right] + \lambda_2 \lambda_6 \left[ K_{13} - \frac{K_6 K_{16}}{sm} \right] \right]$$

and

$$P_{22} = \frac{1}{sm-1} \left[ \lambda_3^2 \left[ K_1 - \frac{K_5^2}{sm} \right] + \lambda_4^2 \left[ K_7 - \frac{K_{15}^2}{sm} \right] + \lambda_5^2 \left[ K_2 - \frac{K_6^2}{sm} \right] + \lambda_6^2 \left[ K_8 - \frac{K_{16}^2}{sm} \right] + 2\lambda_3 \lambda_4 \left[ K_9 - \frac{K_5 K_{15}}{sm} \right] + 2\lambda_3 \lambda_5 \left[ K_3 - \frac{K_5 K_6}{sm} \right] + 2\lambda_3 \lambda_6 \left[ K_{10} - \frac{K_5 K_{16}}{sm} \right] + 2\lambda_4 \lambda_5 \left[ K_{11} - \frac{K_{15} K_6}{sm} \right] + 2\lambda_4 \lambda_6 \left[ K_{12} - \frac{K_{15} K_{16}}{sm} \right] + 2\lambda_5 \lambda_6 \left[ K_{13} - \frac{K_6 K_{16}}{sm} \right] \right]$$



Nineteen sets of K values were obtained for each of the 15 combinations of  $s = 50, 100, 200, \text{ or } 400$  and  $m = 5, 10, 20, \text{ or } 40$ , with the exception of the combination of 400 sires with 40 offspring each. These combinations of number of sires and number of offspring will hereafter be termed "sample size types".

Calculation of the K values initially in the program allowed the use of different sets of  $\lambda$  values with each set of K values. Lambda values were introduced for each of 15 "population types". These were the combinations of herit-

abilities  $\frac{G_{ii}}{P_{ii}}$  for both of the two traits of .2, .5, or .8

and genetic correlations  $\frac{G_{12}}{G_{11}^{1/2} G_{22}^{1/2}}$  of -.5, 0, .2, .5, or .8

with environmental correlations  $\frac{P_{12} - G_{12}}{[P_{11} - G_{11}]^{1/2} [P_{22} - G_{22}]^{1/2}}$

of zero. The phenotypic variances ( $P_{ii}$ ) were chosen to be unity for both traits for all population types.

Although it would have been advantageous to have independent samples for each of the population types for certain statistical reasons, the use of the same sets of K values for a particular sample size type for all sets of values yielded considerably more data for each unit of machine computing time. Due to this arrangement, any differences among population types within a particular sample size type resulted from the influence of the magnitude of the parameters upon the

sampling errors encountered in the sample of 19 runs.

In actual practice, an experimenter is able to detect some estimates which are not reasonable. Three of these situations are negative estimates of additive genetic variance, estimates of the additive genetic variance which are greater than the estimates of the phenotypic variance (heritability estimates greater than one), and estimates of the additive genetic correlation which are greater than one. In an effort to simulate the usual modifications of such unreasonable estimates as are made when sampling errors and not inadequacies of statistical models or of data are decided to be the causes of these seemingly erroneous estimates, the following operations were incorporated into the program:

- (1) If  $\hat{G}_{ii} < 0$ , set  $\hat{G}_{ii} = 0$  and  $\hat{G}_{12} = 0$
- (2) If  $\hat{G}_{ii} > \hat{P}_{ii}$ , set  $\hat{G}_{ii} = \hat{P}_{ii}$
- (3) If  $\hat{G}_{12}^2 > \hat{G}_{11} \hat{G}_{22}$ , set  $\hat{G}_{12} = \pm (\hat{G}_{11} \hat{G}_{22})^{\frac{1}{2}}$

retaining the original algebraic sign for  $\hat{G}_{12}$ . Sets of estimates which were modified by these rules were marked for later identification. The first of the modifications eliminates most of the cases where negative progress is obtained, although the progress might still be small. These modifications could not be incorporated into the equations developed in the preceding sections and will thereby introduce further discrepancies between the results from the equations and the Monte Carlo results.

The method of generating the random deviates was such as to assure normality of distribution except for rounding to the next larger (in absolute value) one-hundredth of a unit. The procedure also assured randomness of successive values generated. However, an examination of a sample of 1000 consecutive variables generated in this manner was conducted in an effort to test these supposed properties. A Chi-square test was made of the correspondence of the observed frequencies of variables to the expected frequencies in one-tenth of a standard deviation grouping. Some coarser grouping of the frequencies of extreme deviates where the frequencies were quite small was made. This test yielded a Chi-square value with 41 degrees of freedom of 42.39, which when compared with tabular Chi-square values, indicates a probability level between .25 and .5 for a value this large or larger. Also, similar Chi-square values were calculated for each of the ten consecutive groups of 100 consecutive variables, all of which indicated satisfactory agreement between observed and expected results. Two tests were made for randomness. The first of these was the "above and below the mean" test, which consisted of comparing the number of "runs" of different lengths of positive values or negative values with the expected number of runs of these lengths for a truly random variable. This test yielded a Chi-square value of 6.74 for six degrees of freedom, which indicates a probability level between .2 and .3. The

other test of randomness was the "up and down" test, which compares the number of runs of different lengths of increasing or decreasing consecutive variables. This test yielded a Chi-square value of 2.27 with four degrees of freedom, which corresponds to a probability level between .5 and .7. Both of these tests of randomness are described in detail in IBM Reference Manual; Random Number Generation and Testing (1959). Although the sample of 1000 was not large enough to rigorously prove or disprove the supposed distribution and randomness, these tests, along with the knowledge of the theory of the method of generation, indicate a satisfactory agreement with the assumed distribution for the purposes of this problem.

After the estimates of the parameters were obtained, the coefficients for an index were calculated in the usual manner. From these coefficients, the true parameters and the estimates of the parameters,  $\Delta H'$ , which is the expected progress of a particular index, and  $\hat{\Delta H}$ , which is estimated progress for the index, were calculated according to the formulae presented earlier.

These values of  $\Delta H'$  and  $\hat{\Delta H}$  were then analyzed for each combination of population type and sample size type to ascertain the effectiveness of the estimation procedure for the purpose of index construction.

## RESULTS

Because of the complexity of the approximate equations developed in an earlier section, it seemed almost impossible to assess the influence of the magnitude of population parameters or of the size of sample (number of sires or number of offspring per sire) by studying the equations. It seemed preferable to calculate numerical values for these functions for various combinations of true parameters (population types) and combinations of number of sires and number of offspring per sire (sample size types) and to study these numerical values and the trends among them. The numerical values also facilitated a comparison of the results from the approximate equations with the results obtained from the Monte Carlo study of the same population types and sample size types.

Two of the functions which seemed to be of primary interest for this study were  $\frac{\Delta H - E[\Delta H']}{\Delta H}$  and  $\frac{E^{\frac{1}{2}}[(\Delta H' - \Delta H)^2]}{\Delta H}$ . These two functions indicate the closeness of the distribution of  $\Delta H'$  values to the corresponding  $\Delta H$  values. The first,  $\frac{\Delta H - E[\Delta H']}{\Delta H}$  is the mean decrease in progress expressed as a fraction of  $\Delta H$ , in other words, the fractional decrease in progress. The second of these functions,  $\frac{E^{\frac{1}{2}}[(\Delta H' - \Delta H)^2]}{\Delta H}$ , is the square root of the mean squared decrease in progress

expressed as a fraction of  $\Delta H$ . Both of these functions are expressed as fractions of  $\Delta H$  in order to facilitate comparisons between different population types with differing  $\Delta H$  values. The second function will always be somewhat larger than the first due to the variation among  $\Delta H'$  values for a particular population type and sample size type. This may be seen by examining the relation,  $E^{\frac{1}{2}}[(\Delta H' - \Delta H)^2] = \left[ [\Delta H - E[\Delta H']]^2 + V[\Delta H'] \right]^{\frac{1}{2}}$ .

These two functions are both measures of the influence of the errors of estimation upon the progress from selection. The second of these functions gives stronger weight, relative to the first, to the  $\Delta H'$  values which represent extreme decreases from the maximum attainable progress. For example, consider two  $\Delta H'$  values, the second representing twice the decrease in progress that the first represents. To the mean squared decrease in progress, the second value will make a contribution four times that of the first, while the contribution to the mean decrease will be only twice as great. So, in making comparisons among different sample size types where the smaller functions are deemed desirable, comparisons among the second functions,  $\frac{E^{\frac{1}{2}}[(\Delta H' - \Delta H)^2]}{\Delta H}$ , would put increased emphasis against the more extreme values of  $\Delta H'$  than when the first function,  $\frac{\Delta H - E[\Delta H']}{\Delta H}$ , is used as a basis of comparison.

Two other functions are of interest in that they are

indicative of the closeness of relationships between the  $\hat{\Delta H}$  values and the corresponding  $\Delta H'$  values. Since a  $\hat{\Delta H}$  value is an estimate of the progress from selection for a particular calculated index, it is more realistic to consider it as an estimate of the corresponding  $\Delta H'$  value than as an estimate of the  $\Delta H$  value for that population type. The two functions which are of interest in indicating the closeness of relationship between these two variables are  $E[\hat{\Delta H}] - E[\Delta H']$ , the mean difference between the two variables, and  $E[(\hat{\Delta H} - \Delta H')^2]$ , the mean squared difference between them. The mean difference between  $\hat{\Delta H}$  and  $\Delta H'$  is the amount by which an estimate of progress tends to over- or under-estimate the progress for that index. The mean squared difference is the measure of the accuracy of this estimation. Of course, the magnitude of the mean squared difference is somewhat dependent upon the magnitude of the mean difference because  $E[(\hat{\Delta H} - \Delta H')^2] = [E(\hat{\Delta H}) - E(\Delta H')]^2 + V(\hat{\Delta H}) + V(\Delta H') - 2 \text{Cov}(\hat{\Delta H}, \Delta H')$ . These two measures of the bias and accuracy in estimating progress need to be jointly evaluated for any sample size type and population type to assess the effectiveness of progress estimation for that sample size type. It did not seem as meaningful to code these values by dividing by  $\Delta H$  or  $(\Delta H)^2$  as was done in the other two functions, although almost certainly these latter two values are somewhat related to the magnitude of the  $\Delta H$  value for the population type which they represent.

The numerical values for these four functions as developed from the approximate equations for each combination of 15 different population and 15 different sample size types are presented in the tables on the following pages. The latter two functions have been divided by the constant  $\bar{i}_s$ , the selection differential in standard units, to facilitate presentation. Of course, the  $\bar{i}_s$  values cancel out of the first two functions. Also presented are the corresponding estimates of these functions obtained from the 19 Monte Carlo observations made at each of these combinations of sample size type and population type. These four functions were estimated as follows:

$$\frac{1}{\Delta H} \left[ \Delta H - \frac{\sum \Delta H'}{19} \right] \text{ estimates } \frac{\Delta H - E[\Delta H']}{\Delta H},$$

$$\frac{1}{\Delta H} \left[ \frac{\sum (\Delta H' - \Delta H)^2}{19} \right]^{\frac{1}{2}} \text{ estimates } \frac{E^{\frac{1}{2}}[(\Delta H' - \Delta H)^2]}{\Delta H},$$

$$\frac{1}{19} [\sum \hat{\Delta H} - \sum \Delta H'] \text{ estimates } E[\hat{\Delta H}] - E[\Delta H'],$$

$$\text{and } \frac{1}{19} [\sum (\hat{\Delta H} - \Delta H')^2] \text{ estimates } E[(\hat{\Delta H} - \Delta H')^2],$$

where the summations extend over the 19 Monte Carlo observations. It would be well to recall that the technique of carrying out the Monte Carlo phase of this study was such that a particular function for a particular sample size type shall be influenced by the same sampling errors for all population types.



The population types which are presented in these tables all have phenotypic variances,  $P_{11}$  and  $P_{22}$ , which are 1.0 and have an environmental correlation,  $r_{E_1E_2} =$

$$\frac{P_{12} - G_{12}}{(P_{11} - G_{11})^{\frac{1}{2}}(P_{22} - G_{22})^{\frac{1}{2}}}, \text{ which is zero. Represented are}$$

all the combinations where both  $G_{11}$  and  $G_{22}$  assume values of

either .2, .5, or .8 and  $r_{G_1G_2} = \frac{G_{12}}{G_{11}^{\frac{1}{2}}G_{22}^{\frac{1}{2}}}$  assumes values of

either -.5, zero, .2, .5, or .8.

Table 1a. Functions of  $\hat{\Delta H}$ ,  $\Delta H'$ , and  $\Delta H = .1491 \bar{i}_s$  when  $G_{11} = G_{22} = .2$ ,  $r_{G_1 G_2} = -.5$ ,  $r_{E_1 E_2} = 0$  and  $P_{11} = P_{22} = 1.0$  for various combinations of  $s$  and  $m^a$

Sample size $s$	type $m$	$\Delta H - E[\Delta H']$	$E^2[(\Delta H' - \Delta H)^2]$	$E[\hat{\Delta H}] - E[\Delta H']$	$E[(\hat{\Delta H} - \Delta H')^2]$
		$\Delta H$	$\Delta H$	$\bar{i}_s$	$\bar{i}_s^2$
50	5	.7758	1.0008	.3119	.0975
		.3353	.4625	.0745	.0245
50	10	.4597	.5838	.1336	.0192
		.2173	.3126	.1166	.0236
50	20	.1936	.2415	.0522	.0048
		.2180	.3100	.0034	.0030
50	40	.0869	.1046	.0237	.0023
		.0932	.1412	.0047	.0027
100	5	.6152	.7891	.1979	.0399
		.2374	.3782	.0986	.0301
100	10	.2556	.3233	.0688	.0068
		.2361	.3735	.0275	.0072
100	20	.0864	.1058	.0233	.0024
		.0456	.0740	.0110	.0034
100	40	.0382	.0442	.0106	.0013
		.0516	.0761	-.0004	.0025
200	5	.4063	.5187	.1139	.0146
		.2066	.3274	.0405	.0100
200	10	.1180	.1479	.0313	.0031
		.0382	.0679	.0008	.0064
200	20	.0374	.0442	.0103	.0014
		.0282	.0449	-.0072	.0011
200	40	.0173	.0192	.0049	.0007
		.0141	.0236	-.0089	.0011
400	5	.2137	.2722	.0565	.0055
		.2287	.3659	.0232	.0048
400	10	.0502	.0614	.0136	.0018
		.0456	.0700	.0095	.0031
400	20	.0168	.0189	.0047	.0007
		.0134	.0213	-.0215	.0017

<sup>a</sup>The upper value of each pair is derived from the approximate equation while the lower value is the corresponding Monte Carlo estimate.

Table 1b. Functions of  $\hat{\Delta H}$ ,  $\Delta H'$ , and  $\Delta H = .2828 \bar{I}_S$  when  $G_{11} = G_{22} = .2$ ,  $r_{G_1 G_2} = 0$ ,  $r_{E_1 E_2} = 0$  and  $P_{11} = P_{22} = 1.0$  for various combinations of  $s$  and  $m^a$

Sample size $s$	type $m$	$\frac{\Delta H - E[\Delta H']}{\Delta H}$	$\frac{E^{\frac{1}{2}}[(\Delta H' - \Delta H)^2]}{\Delta H}$	$\frac{E[\hat{\Delta H}] - E[\Delta H']}{\bar{I}_S}$	$\frac{E[(\hat{\Delta H} - \Delta H')^2]}{\bar{I}_S^2}$
50	5	.5162 .1878	.6617 .2835	.2859 .0230	.0861 .0374
50	10	.1854 .1344	.2368 .2067	.0913 .0908	.0173 .0305
50	20	.0645 .1033	.0813 .2109	.0321 -.0469	.0084 .0065
50	40	.0303 .0311	.0371 .0454	.0154 -.0321	.0054 .0080
100	5	.2890 .1096	.3698 .1989	.1456 .0645	.0291 .0395
100	10	.0778 .0477	.0987 .0879	.0387 -.0184	.0092 .0149
100	20	.0260 .0127	.0318 .0250	.0134 -.0030	.0049 .0096
100	40	.0126 .0173	.0148 .0317	.0067 -.0216	.0030 .0076
200	5	.1331 .0492	.1703 .0752	.0658 -.0083	.0130 .0145
200	10	.0318 .0103	.0392 .0170	.0163 -.0188	.0054 .0119
200	20	.0111 .0095	.0129 .0156	.0059 -.0260	.0027 .0040
200	40	.0056 .0050	.0062 .0085	.0030 -.0212	.0016 .0040
400	5	.0550 .0962	.0694 .1951	.0277 -.0152	.0075 .0090
400	10	.0133 .0138	.0156 .0189	.0071 -.0121	.0030 .0069
400	20	.0049 .0039	.0055 .0057	.0027 -.0399	.0014 .0052

<sup>a</sup>The upper value of each pair is derived from the approximate equation while the lower value is the corresponding Monte Carlo estimate.

Table 1c. Functions of  $\hat{\Delta H}$ ,  $\Delta H'$ , and  $\Delta H = .3328 \bar{I}_s$  when  $G_{11} = G_{22} = .2$ ,  $r_{G_1 G_2} = .2$ ,  $r_{E_1 E_2} = 0$  and  $P_{11} = P_{22} = 1.0$  for various combinations of  $s$  and  $m^a$

Sample size $s$	type $m$	$\frac{\Delta H - E[\Delta H']}{\Delta H}$	$\frac{E[\Delta H' - \Delta H]^2}{\Delta H}$	$\frac{E[\hat{\Delta H}] - E[\Delta H']}{\bar{I}_s}$	$\frac{E[(\hat{\Delta H} - \Delta H')^2]}{\bar{I}_s^2}$
50	5	.4350 .2236	.5575 .3604	.2698 .0243	.0809 .0399
50	10	.1392 .1094	.1785 .1804	.0798 .0732	.0191 .0335
50	20	.0361 .0724	.0496 .1441	.0239 -.0673	.0104 .0090
50	40	.0225 .0216	.0280 .0321	.0133 -.0479	.0070 .0111
100	5	.2202 .0832	.2823 .1637	.1281 .0551	.0286 .0450
100	10	.0561 .0358	.0713 .0668	.0328 -.0275	.0110 .0177
100	20	.0190 .0087	.0233 .0163	.0114 -.0073	.0060 .0130
100	40	.0093 .0120	.0110 .0236	.0057 -.0293	.0038 .0107
200	5	.0953 .0376	.1218 .0568	.0554 -.0210	.0146 .0160
200	10	.0226 .0072	.0278 .0111	.0137 -.0257	.0064 .0144
200	20	.0081 .0072	.0094 .0116	.0050 -.0338	.0033 .0057
200	40	.0041 .0036	.0046 .0058	.0026 -.0254	.0020 .0058
400	5	.0386 .0619	.0484 .1216	.0231 -.0302	.0086 .0123
400	10	.0096 .0102	.0112 .0140	.0060 -.0206	.0035 .0090
400	20	.0036 .0030	.0040 .0042	.0023 -.0469	.0017 .0070

<sup>a</sup>The upper value of each pair is derived from the approximate equation while the lower value is the corresponding Monte Carlo estimate.

Table 1d. Functions of  $\hat{\Delta H}$ ,  $\Delta H'$ , and  $\Delta H = .4045 \bar{I}_s$  when  $G_{11} = G_{22} = .2$ ,  $r_{G_1 G_2} = .5$ ,  $r_{E_1 E_2} = 0$  and  $P_{11} = P_{22} = 1.0$  for various combinations of  $s$  and  $m^a$

Sample size $s$	type $m$	$\frac{\Delta H - E[\Delta H']}{\Delta H}$	$\frac{E^{\frac{1}{2}}[(\Delta H' - \Delta H)^2]}{\Delta H}$	$\frac{E[\hat{\Delta H}] - E[\Delta H']}{\bar{I}_s}$	$\frac{E[(\hat{\Delta H} - \Delta H')^2]}{\bar{I}_s^2}$
50	5	.3406	.4370	.2456	.0759
		.2054	.3510	.0109	.0470
50	10	.0967	.1250	.0667	.0232
		.0769	.1367	.0415	.0387
50	20	.0325	.0419	.0226	.0138
		.0304	.0576	-.0960	.0180
50	40	.0151	.0194	.0105	.0096
		.0126	.0196	-.0727	.0170
100	5	.1544	.1985	.1077	.0305
		.0559	.1137	.0461	.0518
100	10	.0377	.0481	.0267	.0138
		.0230	.0442	.0422	.0232
100	20	.0128	.0159	.0091	.0078
		.0054	.0084	.0135	.0187
100	40	.0061	.0075	.0043	.0052
		.0064	.0136	.0402	.0168
200	5	.0634	.0809	.0449	.0173
		.0277	.0394	-.0399	.0199
200	10	.0152	.0186	.0111	.0079
		.0044	.0069	-.0366	.0185
200	20	.0055	.0064	.0040	.0042
		.0049	.0076	-.0460	.0087
200	40	.0027	.0031	.0019	.0027
		.0020	.0029	-.0311	.0091
400	5	.0256	.0317	.0187	.0102
		.0307	.0522	-.0473	.0184
400	10	.0066	.0076	.0049	.0042
		.0067	.0099	-.0341	.0128
400	20	.0025	.0028	.0018	.0021
		.0020	.0029	-.0575	.0099

<sup>a</sup>The upper value of each pair is derived from the approximate equation while the lower value is the corresponding Monte Carlo estimate.

Table 1e. Functions of  $\hat{\Delta}H$ ,  $\Delta H'$ , and  $\Delta H = .4727 \bar{I}_s$  when  $G_{11} = G_{22} = .2$ ,  $r_{G_1 G_2} = .8$ ,  $r_{E_1 E_2} = 0$  and  $P_{11} = P_{22} = 1.0$  for various combinations of  $s$  and  $m^a$

Sample size $s$	type $m$	$\frac{\Delta H - E[\Delta H']}{\Delta H}$	$\frac{E^{\frac{1}{2}}[(\Delta H' - \Delta H)^2]}{\Delta H}$	$\frac{E[\hat{\Delta}H] - E[\Delta H']}{\bar{I}_s}$	$\frac{E[(\hat{\Delta}H - \Delta H')^2]}{\bar{I}_s^2}$
50	5	.2719 .1807	.3499 .2856	.2233 -.0142	.0745 .0571
50	10	.0710 .0425	.0926 .0826	.0566 -.0003	.0280 .0493
50	20	.0230 .0184	.0305 .0359	.0182 -.1053	.0174 .0225
50	40	.0101 .0074	.0137 .0119	.0077 -.0990	.0124 .0247
100	5	.1143 .0360	.1474 .0721	.0926 .0330	.0342 .0565
100	10	.0271 .0146	.0348 .0288	.0222 -.0673	.0167 .0327
100	20	.0089 .0044	.0113 .0067	.0072 -.0240	.0097 .0235
100	40	.0040 .0036	.0051 .0061	.0030 -.0504	.0066 .0242
200	5	.0457 .0248	.0582 .0334	.0378 -.0670	.0202 .0272
200	10	.0109 .0032	.0134 .0053	.0092 -.0502	.0093 .0228
200	20	.0038 .0032	.0045 .0046	.0031 -.0614	.0051 .0123
200	40	.0018 .0013	.0021 .0016	.0013 -.0365	.0034 .0136
400	5	.0185 .0239	.0228 .0431	.0158 -.0657	.0117 .0248
400	10	.0048 .0047	.0055 .0083	.0041 -.0512	.0050 .0183
400	20	.0017 .0013	.0019 .0020	.0014 -.0689	.0026 .0130

<sup>a</sup>The upper value of each pair is derived from the approximate equation while the lower value is the corresponding Monte Carlo estimate.

Table 1f. Functions of  $\hat{\Delta H}$ ,  $\Delta H'$ , and  $\Delta H = .4082 \bar{I}_S$  when  $G_{11} = G_{22} = .5$ ,  $r_{G_1 G_2} = -.5$ ,  $r_{E_1 E_2} = 0$  and  $P_{11} = P_{22} = 1.0$  for various combinations of  $s$  and  $m^a$

Sample size $s$	type $m$	$\frac{\Delta H - E[\Delta H']}{\Delta H}$	$\frac{E^{\frac{1}{2}}[(\Delta H' - \Delta H)^2]}{\Delta H}$	$\frac{E[\hat{\Delta H}] - E[\Delta H']}{\bar{I}_S}$	$\frac{E[(\hat{\Delta H} - \Delta H')^2]}{\bar{I}_S^2}$
50	5	.2685	.3395	.1948	.0545
		.1744	.2925	.0213	.0311
50	10	.0901	.1110	.0652	.0199
		.1124	.1907	.0898	.0427
50	20	.0433	.0511	.0321	.0120
		.0590	.1071	-.0712	.0135
50	40	.0283	.0322	.0213	.0086
		.0363	.0541	-.0233	.0146
100	5	.1144	.1438	.0820	.0239
		.0674	.1163	.0163	.0391
100	10	.0376	.0448	.0279	.0114
		.0343	.0664	-.0221	.0159
100	20	.0188	.0212	.0143	.0066
		.0145	.0260	-.0149	.0157
100	40	.0127	.0138	.0097	.0046
		.0211	.0335	-.0251	.0113
200	5	.0474	.0583	.0348	.0137
		.0265	.0470	-.0254	.0171
200	10	.0163	.0185	.0124	.0063
		.0096	.0171	-.0236	.0173
200	20	.0086	.0093	.0066	.0035
		.0088	.0125	-.0309	.0068
200	40	.0059	.0062	.0046	.0024
		.0061	.0102	-.0350	.0072
400	5	.0200	.0236	.0152	.0078
		.0201	.0372	-.0520	.0186
400	10	.0074	.0080	.0058	.0034
		.0076	.0113	-.0179	.0098
400	20	.0040	.0042	.0032	.0018
		.0027	.0047	-.0506	.0100

<sup>a</sup>The upper value of each pair is derived from the approximate equation while the lower value is the corresponding Monte Carlo estimate.

Table 1g. Functions of  $\hat{\Delta H}$ ,  $\Delta H'$ , and  $\Delta H = .7071 \bar{I}_S$  when  $G_{11} = G_{22} = .5$ ,  $r_{G_1 G_2} = 0$ ,  $r_{E_1 E_2} = 0$  and  $P_{11} = P_{22} = 1.0$  for various combinations of  $s$  and  $m^a$

Sample size $s$	type $m$	$\Delta H - E[\Delta H']$ $\Delta H$	$E^{\frac{1}{2}}[(\Delta H' - \Delta H)^2]$ $\Delta H$	$E[\hat{\Delta H}] - E[\Delta H']$ $\bar{I}_S$	$E[(\hat{\Delta H} - \Delta H')^2]$ $\bar{I}_S^2$
50	5	.1194 .0649	.1510 .1110	.1449 -.0636	.0781 .0723
50	10	.0389 .0499	.0475 .0807	.0477 .0507	.0417 .0790
50	20	.0202 .0294	.0237 .0458	.0248 -.1566	.0283 .0376
50	40	.0139 .0181	.0159 .0263	.0170 -.0892	.0221 .0409
100	5	.0447 .0154	.0553 .0217	.0554 -.0217	.0437 .0819
100	10	.0161 .0164	.0188 .0339	.0201 -.0572	.0234 .0375
100	20	.0087 .0044	.0098 .0088	.0108 -.0391	.0152 .0485
100	40	.0062 .0096	.0067 .0188	.0076 -.0612	.0116 .0369
200	5	.0185 .0126	.0219 .0226	.0235 -.0674	.0249 .0326
200	10	.0071 .0041	.0079 .0058	.0090 -.0564	.0125 .0412
200	20	.0040 .0041	.0043 .0064	.0050 -.0704	.0079 .0240
200	40	.0029 .0033	.0030 .0057	.0036 -.0590	.0060 .0231
400	5	.0081 .0092	.0091 .0178	.0105 -.0870	.0134 .0383
400	10	.0033 .0040	.0035 .0059	.0042 -.0539	.0065 .0280
400	20	.0019 .0014	.0020 .0021	.0024 -.0871	.0040 .0287

<sup>a</sup>The upper value of each pair is derived from the approximate equation while the lower value is the corresponding Monte Carlo estimate.



Table 1h. Functions of  $\hat{\Delta}H$ ,  $\Delta H'$ , and  $\Delta H = .8091 \bar{I}_S$  when  $G_{11} = G_{22} = .5$ ,  $r_{G_1 G_2} = .2$ ,  $r_{E_1 E_2} = 0$  and  $P_{11} = P_{22} = 1.0$  for various combinations of  $s$  and  $m^a$

Sample size $s$	type $m$	$\Delta H - E[\Delta H']$ $\Delta H$	$E^{\frac{1}{2}}[(\Delta H' - \Delta H)^2]$ $\Delta H$	$E[\hat{\Delta}H] - E[\Delta H']$ $\bar{I}_S$	$E[(\hat{\Delta}H - \Delta H')^2]$ $\bar{I}_S^2$
50	5	.0975 .0439	.1233 .0912	.1346 -.0890	.0919 .1016
50	10	.0316 .0389	.0387 .0613	.0438 .0346	.0512 .0934
50	20	.0164 .0215	.0194 .0356	.0225 -.1856	.0354 .0528
50	40	.0113 .0145	.0130 .0215	.0152 -.1177	.0281 .0543
100	5	.0360 .0133	.0444 .0175	.0508 -.0240	.0520 .0941
100	10	.0131 .0129	.0153 .0259	.0184 -.0721	.0284 .0501
100	20	.0071 .0028	.0080 .0054	.0098 -.0477	.0188 .0641
100	40	.0050 .0070	.0055 .0146	.0067 -.0751	.0147 .0513
200	5	.0150 .0104	.0177 .0182	.0216 -.0829	.0293 .0398
200	10	.0058 .0031	.0065 .0041	.0082 -.0687	.0150 .0516
200	20	.0033 .0035	.0035 .0055	.0045 -.0862	.0097 .0327
200	40	.0024 .0026	.0025 .0044	.0032 -.0669	.0075 .0315
400	5	.0067 .0074	.0074 .0141	.0097 -.0979	.0156 .0467
400	10	.0027 .0033	.0029 .0052	.0039 -.0689	.0077 .0371
400	20	.0016 .0014	.0017 .0018	.0022 -.1003	.0049 .0371

<sup>a</sup>The upper value of each pair is derived from the approximate equation while the lower value is the corresponding Monte Carlo estimate.

Table 1i. Functions of  $\hat{\Delta}H$ ,  $\Delta H'$ , and  $\Delta H = .9487 \bar{I}_S$  when  $G_{11} = G_{22} = .5$ ,  $r_{G_1 G_2} = .5$ ,  $r_{E_1 E_2} = 0$  and  $P_{11} = P_{22} = 1.0$  for various combinations of  $s$  and  $m^a$

Sample size $s$	type $m$	$\frac{\Delta H - E[\Delta H']}{\Delta H}$	$E^{\frac{1}{2}}[(\Delta H' - \Delta H)^2]$ $\Delta H$	$\frac{E[\hat{\Delta}H] - E[\Delta H']}{\bar{I}_S}$	$\frac{E[(\hat{\Delta}H - \Delta H')^2]}{\bar{I}_S^2}$
50	5	.0690 .0387	.0888 .0853	.1157 -.1053	.1145 .1267
50	10	.0242 .0246	.0297 .0370	.0383 .0046	.0657 .1108
50	20	.0122 .0138	.0146 .0270	.0186 -.2170	.0464 .0741
50	40	.0082 .0098	.0096 .0140	.0120 -.1629	.0376 .0786
100	5	.0279 .0119	.0343 .0154	.0457 -.0223	.0644 .1075
100	10	.0099 .0083	.0116 .0154	.0159 -.0959	.0360 .0737
100	20	.0052 .0017	.0059 .0024	.0079 -.0605	.0245 .0875
100	40	.0036 .0037	.0040 .0082	.0051 -.0948	.0195 .0774
200	5	.0117 .0075	.0137 .0122	.0195 -.1059	.0357 .0516
200	10	.0044 .0021	.0049 .0027	.0071 -.0863	.0189 .0671
200	20	.0024 .0026	.0026 .0041	.0036 -.1097	.0126 .0470
200	40	.0017 .0015	.0018 .0024	.0028 -.0775	.0091 .0466
400	5	.0052 .0052	.0058 .0094	.0088 -.1108	.0189 .0598
400	10	.0021 .0025	.0022 .0042	.0033 -.0929	.0097 .0530
400	20	.0012 .0011	.0012 .0015	.0017 -.1195	.0064 .0500

<sup>a</sup>The upper value of each pair is derived from the approximate equation while the lower value is the corresponding Monte Carlo estimate.

Table 1j. Functions of  $\hat{\Delta}H$ ,  $\Delta H'$ , and  $\Delta H = 1.0757 \bar{I}_S$  when  $G_{11} = G_{22} = .5$ ,  $r_{G_1 G_2} = .8$ ,  $r_{E_1 E_2} = 0$  and  $P_{11} = P_{22} = 1.0$  for various combinations of  $s$  and  $m^a$

Sample size $s$	type $m$	$\frac{\Delta H - E[\Delta H']}{\Delta H}$	$E^{\frac{1}{2}}[(\Delta H' - \Delta H)^2]$	$\frac{E[\hat{\Delta}H] - E[\Delta H']}{\bar{I}_S}$	$\frac{E[(\hat{\Delta}H - \Delta H')^2]}{\bar{I}_S^2}$
50	5	.0619 .0247	.0791 .0351	.1117 -.1351	.1369 .1733
50	10	.0183 .0123	.0230 .0184	.0316 -.0306	.0808 .1402
50	20	.0084 .0090	.0104 .0184	.0132 -.2274	.0580 .0867
50	40	.0051 .0051	.0063 .0078	.0071 -.2121	.0476 .1110
100	5	.0223 .0097	.0276 .0132	.0410 -.0190	.0768 .1117
100	10	.0074 .0046	.0088 .0076	.0128 -.1222	.0438 .1009
100	20	.0035 .0017	.0041 .0025	.0053 -.0761	.0303 .1066
100	40	.0021 .0016	.0025 .0027	.0028 -.1132	.0245 .1080
200	5	.0093 .0054	.0109 .0075	.0174 -.1306	.0421 .0655
200	10	.0032 .0016	.0036 .0021	.0056 -.1046	.0228 .0806
200	20	.0015 .0016	.0017 .0021	.0023 -.1343	.0155 .0624
200	40	.0009 .0006	.0010 .0008	.0012 -.0877	.0124 .0653
400	5	.0041 .0036	.0046 .0065	.0079 -.1184	.0221 .0727
400	10	.0015 .0017	.0016 .0030	.0026 -.1207	.0116 .0723
400	20	.0007 .0007	.0008 .0010	.0011 -.1397	.0078 .0627

<sup>a</sup>The upper value of each pair is derived from the approximate equation while the lower value is the corresponding Monte Carlo estimate.

Table 1k. Functions of  $\hat{\Delta}H$ ,  $\Delta H'$ , and  $\Delta H = .7303 \bar{i}_s$  when  $G_{11} = G_{22} = .8$ ,  $r_{G_1 G_2} = -.5$ ,  $r_{E_1 E_2} = 0$  and  $P_{11} = P_{22} = 1.0$  for various combinations of  $s$  and  $m^a$

Sample size $s$	type $m$	$\frac{\Delta H - E[\Delta H']}{\Delta H}$	$\frac{E^{\frac{1}{2}}[(\Delta H' - \Delta H)^2]}{\Delta H}$	$\frac{E[\hat{\Delta}H] - E[\Delta H']}{\bar{i}_s}$	$\frac{E[(\hat{\Delta}H - \Delta H')^2]}{\bar{i}_s^2}$
50	5	.0700 .0793	.0862 .1797	.0880 -.1235	.0609 .0556
50	10	.0276 .0338	.0321 .0628	.0348 -.0329	.0354 .0912
50	20	.0172 .0251	.0193 .0506	.0216 -.1663	.0255 .0517
50	40	.0134 .0175	.0147 .0261	.0168 -.0755	.0211 .0389
100	5	.0261 .0142	.0309 .0227	.0402 -.0606	.0238 .0662
100	10	.0119 .0140	.0132 .0277	.0152 -.0565	.0193 .0291
100	20	.0077 .0066	.0083 .0117	.0097 -.0484	.0135 .0443
100	40	.0061 .0107	.0065 .0178	.0077 -.0611	.0110 .0326
200	5	.0113 .0086	.0127 .0153	.0147 -.0729	.0184 .0321
200	10	.0055 .0047	.0058 .0076	.0070 -.0559	.0101 .0379
200	20	.0036 .0041	.0038 .0055	.0046 -.0640	.0069 .0231
200	40	.0029 .0033	.0030 .0055	.0037 -.0676	.0056 .0230
400	5	.0052 .0052	.0056 .0113	.0068 -.0874	.0097 .0374
400	10	.0026 .0026	.0027 .0040	.0034 -.0487	.0051 .0262
400	20	.0018 .0012	.0018 .0020	.0022 -.0862	.0035 .0294

<sup>a</sup>The upper value of each pair is derived from the approximate equation while the lower value is the corresponding Monte Carlo estimate.

Table 11. Functions of  $\hat{\Delta}H$ ,  $\Delta H'$ , and  $\Delta H = 1.1314 \bar{I}_S$  when  $G_{11} = G_{22} = .8$ ,  $r_{G_1 G_2} = 0$ ,  $r_{E_1 E_2} = 0$  and  $P_{11} = P_{22} = 1.0$  for various combinations of  $s$  and  $m^a$

Sample size $s$	type $m$	$\frac{\Delta H - E[\Delta H']}{\Delta H}$	$\frac{E^{\frac{1}{2}}[(\Delta H' - \Delta H)^2]}{\Delta H}$	$\frac{E[\hat{\Delta}H] - E[\Delta H']}{\bar{I}_S}$	$\frac{E[(\hat{\Delta}H - \Delta H')^2]}{\bar{I}_S^2}$
50	5	.0455 .0339	.0554 .0947	.0873 -.1815	.1223 .1123
50	10	.0187 .0245	.0214 .0504	.0348 -.0552	.0744 .1346
50	20	.0120 .0177	.0133 .0270	.0218 -.2519	.0560 .0986
50	40	.0095 .0117	.0103 .0174	.0169 -.1701	.0476 .0942
100	5	.0170 .0060	.0197 .0081	.0329 -.0877	.0658 .1179
100	10	.0082 .0093	.0089 .0189	.0152 -.0975	.0396 .0762
100	20	.0054 .0026	.0058 .0051	.0097 -.0769	.0291 .1146
100	40	.0044 .0070	.0046 .0140	.0076 -.1031	.0245 .0888
200	5	.0075 .0063	.0083 .0108	.0146 -.1230	.0351 .0645
200	10	.0038 .0027	.0040 .0038	.0070 -.0970	.0205 .0865
200	20	.0026 .0028	.0027 .0043	.0046 -.1171	.0149 .0619
200	40	.0021 .0027	.0022 .0045	.0036 -.0976	.0124 .0584
400	5	.0035 .0040	.0037 .0082	.0068 -.1317	.0182 .0778
	10	.0018 .0022	.0019 .0036	.0034 -.0954	.0104 .0657
400	20	.0013 .0009	.0013 .0012	.0022 -.1342	.0075 .0705

<sup>a</sup>The upper value of each pair is derived from the approximate equation while the lower value is the corresponding Monte Carlo estimate.

Table 1m. Functions of  $\hat{\Delta}H$ ,  $\Delta H'$ , and  $\Delta H = 1.2605 \bar{I}_s$  when  $G_{11} = G_{22} = .8$ ,  $r_{G_1 G_2} = .2$ ,  $r_{E_1 E_2} = 0$  and  $P_{11} = P_{22} = 1.0$  for various combinations of  $s$  and  $m^a$

Sample size $s$	type $m$	$\frac{\Delta H - E[\Delta H']}{\Delta H}$	$\frac{E^{\frac{1}{2}}[(\Delta H' - \Delta H)^2]}{\Delta H}$	$\frac{E[\hat{\Delta}H] - E[\Delta H']}{\bar{I}_s}$	$\frac{E[(\hat{\Delta}H - \Delta H')^2]}{\bar{I}_s^2}$
50	5	.0415	.0505	.0885	.1471
		.0265	.0709	-.1968	.1379
50	10	.0170	.0195	.0349	.0902
		.0214	.0391	-.0651	.1489
50	20	.0109	.0121	.0216	.0684
		.0155	.0264	-.2803	.1192
50	40	.0086	.0093	.0166	.0585
		.0110	.0168	-.1993	.1201
100	5	.0155	.0179	.0333	.0786
		.0059	.0077	-.0920	.1298
100	10	.0074	.0081	.0152	.0479
		.0082	.0161	-.1136	.0990
100	20	.0049	.0052	.0096	.0355
		.0017	.0032	-.0873	.1447
100	40	.0039	.0041	.0075	.0300
		.0056	.0118	-.1177	.1173
200	5	.0069	.0075	.0148	.0418
		.0056	.0101	-.1391	.0784
200	10	.0034	.0036	.0070	.0247
		.0022	.0030	-.1112	.1064
200	20	.0023	.0023	.0045	.0181
		.0025	.0041	-.1377	.0797
200	40	.0019	.0019	.0035	.0152
		.0022	.0039	-.1064	.0752
400	5	.0032	.0034	.0069	.0216
		.0035	.0071	-.1446	.0937
400	10	.0016	.0017	.0033	.0125
		.0021	.0037	-.1145	.0835
400	20	.0011	.0011	.0022	.0091
		.0009	.0013	-.1504	.0870

<sup>a</sup>The upper value of each pair is derived from the approximate equation while the lower value is the corresponding Monte Carlo estimate.

Table 1n. Functions of  $\hat{\Delta}H$ ,  $\Delta H'$ , and  $\Delta H = 1.4342 \bar{I}_S$  when  $G_{11} = G_{22} = .8$ ,  $r_{G_1 G_2} = .5$ ,  $r_{E_1 E_2} = 0$  and  $P_{11} = P_{22} = 1.0$  for various combinations of  $s$  and  $m^a$

Sample size $s$	type $m$	$\frac{\Delta H - E[\Delta H']}{\Delta H}$	$\frac{E^{\frac{1}{2}}[(\Delta H' - \Delta H)^2]}{\Delta H}$	$\frac{E[\hat{\Delta}H] - E[\Delta H']}{\bar{I}_S}$	$\frac{E[(\hat{\Delta}H - \Delta H')^2]}{\bar{I}_S^2}$
50	5	.0373	.0453	.0894	.1844
		.0199	.0461	-.2116	.1763
50	10	.0149	.0171	.0340	.1141
		.0155	.0279	-.0857	.1739
50	20	.0094	.0104	.0203	.0872
		.0127	.0243	-.3142	.1516
50	40	.0073	.0080	.0151	.0750
		.0100	.0141	-.2432	.1655
100	5	.0139	.0161	.0334	.0978
		.0066	.0108	-.0937	.1393
100	10	.0065	.0071	.0146	.0602
		.0057	.0095	-.1438	.1436
100	20	.0042	.0045	.0089	.0451
		.0010	.0015	-.1069	.1845
100	40	.0033	.0035	.0067	.0384
		.0033	.0073	-.1421	.1680
200	5	.0062	.0068	.0149	.0518
		.0043	.0071	-.1667	.1023
200	10	.0030	.0032	.0067	.0310
		.0017	.0022	-.1325	.1342
200	20	.0020	.0021	.0042	.0230
		.0023	.0038	-.1683	.1078
200	40	.0016	.0016	.0032	.0194
		.0015	.0027	-.1183	.1041
400	5	.0029	.0030	.0070	.0267
		.0027	.0050	-.1592	.1175
400	10	.0014	.0015	.0032	.0157
		.0019	.0036	-.1451	.1128
400	20	.0009	.0010	.0020	.0116
		.0010	.0014	-.1736	.1111

<sup>a</sup>The upper value of each pair is derived from the approximate equation while the lower value is the corresponding Monte Carlo estimate.

Table 10. Functions of  $\hat{\Delta}H$ ,  $\Delta H'$ , and  $\Delta H = 1.5902 \bar{i}_s$  when  $G_{11} = G_{22} = .8$ ,  $r_{G_1 G_2} = .8$ ,  $r_{E_1 E_2} = 0$  and  $P_{11} = P_{22} = 1.0$  for various combinations of  $s$  and  $m^a$

Sample size $s$	type $m$	$\Delta H - E[\Delta H']$ $\Delta H$	$E^{\frac{1}{2}}[(\Delta H' - \Delta H)^2]$ $\Delta H$	$E[\hat{\Delta}H] - E[\Delta H']$ $\bar{i}_s$	$E[(\hat{\Delta}H - \Delta H')^2]$ $\bar{i}_s^2$
50	5	.0330 .0156	.0405 .0293	.0862 -.2173	.2222 .2166
50	10	.0103 .0074	.0126 .0128	.0247 -.1125	.1433 .0307
50	20	.0072 .0101	.0081 .0191	.0153 -.3226	.1065 .1708
50	40	.0053 .0064	.0059 .0093	.0102 -.3028	.0919 .2255
100	5	.0123 .0063	.0143 .0112	.0319 -.0928	.1172 .1475
100	10	.0053 .0035	.0059 .0050	.0123 -.1792	.0728 .1930
100	20	.0032 .0011	.0035 .0018	.0065 -.1295	.0548 .2156
100	40	.0024 .0014	.0025 .0023	.0043 -.1648	.0469 .2264
200	5	.0054 .0028	.0060 .0037	.0141 -.1950	.0618 .1291
200	10	.0025 .0016	.0026 .0022	.0056 -.1544	.0373 .1577
200	20	.0015 .0018	.0016 .0026	.0030 -.1998	.0279 .1371
200	40	.0011 .0007	.0012 .0011	.0020 -.1299	.0237 .1390
400	5	.0026 .0019	.0027 .0034	.0066 -.1660	.0318 .1399
400	10	.0012 .0015	.0012 .0029	.0026 -.1806	.0189 .1471
400	20	.0007 .0008	.0008 .0012	.0014 -.1976	.0140 .1340

<sup>a</sup> The upper value of each pair is derived from approximate equation while the lower value is the corresponding Monte Carlo estimate.



## DISCUSSION

Several interesting relationships may be studied and evaluated by inspection of the foregoing tables. One of the fundamental points of interest for this study and its possible future extensions is the relationship between the numerical values obtained from the approximate equations and the corresponding estimates of the numerical values obtained from the Monte Carlo phase of this study. Any discrepancy between an equational result and its corresponding Monte Carlo estimate will be the result of one or more of the three following causes: (1) the discrepancy between the equational results and the corresponding true value due to the inaccuracies of the approximation technique used in the development of the equations, (2) the sampling errors involved in the Monte Carlo estimates, or (3) the modifications incorporated into the Monte Carlo procedure for adjusting seemingly erroneous estimates. Since the Monte Carlo estimates are based on only 19 observations, these numerical estimates were expected to be somewhat erratic. However, by studying the trends among these Monte Carlo estimates, one can obtain an indication of the discrepancies between the approximate equations and the Monte Carlo estimates.

In studying trends among these four functions of interest, one should keep in mind the complexity of the relationships

involved and thus realize the difficulty of providing a simple explanation of these trends. A particular  $\Delta H'$  value is determined by the associated  $\hat{b}$  values and by the true parameter values of that population type. The  $\hat{b}$  values are complex functions of the  $\hat{G}_{ij}$  and  $\hat{P}_{ij}$  values, the estimates of the true parameters. For the situations studied here, there are six of these estimates,  $\hat{G}_{11}$ ,  $\hat{G}_{12}$ ,  $\hat{G}_{22}$ ,  $\hat{P}_{11}$ ,  $\hat{P}_{12}$ , and  $\hat{P}_{22}$ . The complex relationship between the  $\hat{b}$  values and the estimates of the parameters is seen from the equations for obtaining the  $\hat{b}$  values, which in matrix notation are  $\hat{b} = \hat{P}^{-1} \hat{G}a$ . A particular  $\Delta H$  value is determined from the same  $\hat{b}$  values and the estimates of the genetic variances and covariances. It is therefore seen that the four functions which were tabulated are dependent upon the true parameters and upon the joint sampling distribution of the six estimates involved. The nature of this joint sampling distribution is rather complex and is determined by the estimation procedure, the magnitude of the true parameters involved, the volume of data which is used, and the distribution of the basic data. The estimation procedure which was considered in this study was the analyses of variance and covariance of paternal-half-sib groups, and the basic data were assumed to be distributed normally. Due to the complexity involved here the trends which can be noted among the numerical values for changes in the true parameters and changes in the sample size types are rather complex and

cannot be easily explained. The purpose of presentation here was to realize the direction of these trends and to gain some appreciation for their magnitude, not to give a detailed explanation of the reasons for their existence.

In all cases the directional trends for the functions presented in the table are the same for the equational and Monte Carlo results when the number of sires ( $s$ ), the number of offspring per sire ( $m$ ), or the magnitude of one of the parameters ( $G_{11}$  and  $G_{22}$  or  $r_{G_1G_2}$ ) is changed. However, certain differences are noticed in the actual numerical values. For the fractional decrease in progress, the first function presented in the table, the equational values tend to be somewhat greater than the corresponding Monte Carlo values when  $s$ ,  $m$ , and/or  $G_{11}$  and  $G_{22}$  are small. This can, in part, be explained by the procedure incorporated into the Monte Carlo routine of setting negative estimates of  $G_{11}$  or  $G_{22}$  equal to zero for the further calculations. This would tend to alleviate some of the more extreme decreases in progress which would be more frequent when  $s$ ,  $m$ , and  $G_{11}$  and  $G_{22}$  are small. Whether this completely explains the noted discrepancies cannot be decided at present.

In comparing the equational and Monte Carlo values for the square root of the mean squared decrease expressed as a fraction of  $\Delta H$ , one notes a seeming tendency for the Monte Carlo estimates to slightly exceed the corresponding

equational value when either  $s$  or  $m$  is large. The previously noted tendency for the equational values for the fractional decrease in progress to exceed the corresponding Monte Carlo results when  $s$ ,  $m$  and/or  $G_{11}$  and  $G_{22}$  are small results in a similar tendency in the second function because of its relationship to the first.

The results from the approximate equations and their Monte Carlo estimates seem to be extremely divergent for the third function, the mean difference between the  $\hat{\Delta H}$  values and their corresponding  $\Delta H'$  values. The equational results tend decidedly to exceed their corresponding Monte Carlo estimates. This decided discrepancy is such that when  $s$ ,  $m$ ,  $G_{11}$  and  $G_{22}$ , or  $r_{G_1 G_2}$  is large, the equational results are all positive, while the Monte Carlo results are predominantly negative. There is no readily apparent explanation for this extreme discrepancy between the results.

For the last function, the mean squared difference between the  $\Delta H$  and the  $\Delta H'$  values, a slight tendency is noted for the equational values to exceed the Monte Carlo values when both  $s$  and  $m$  are small. The discrepancy is larger when  $m$ , the number of offspring per sire, is large. In these cases the equational results tend to be less than the Monte Carlo results. This tendency is especially strong when the values of  $G_{11}$  and  $G_{22}$  and  $r_{G_1 G_2}$  are large.

An over-all consideration of these discrepancies leads

to the conclusion that the equational results are definitely not satisfactory for studying the relationship between the  $\hat{\Delta H}$  and  $\Delta H'$  values. In these functions,  $E[\hat{\Delta H}] - E[\Delta H']$  and  $E[(\hat{\Delta H} - \Delta H')^2]$  the equational results do not seem to be at all valid. However, for these two functions the Monte Carlo results are somewhat erratic when only 19 observations are used. This erratic behavior makes the evaluation of trends and comparisons among sample size types difficult.

For the functions which measure the closeness of the population of  $\Delta H'$  values to their corresponding  $\Delta H$  values, although slight discrepancies seem to exist, the equational values may still be of some value in studying the trends and making comparisons among different sample size types.

As either  $s$ , the number of sires, or  $m$ , the number of offspring per sire, increases, both of the two functions, which are measures of the magnitude of the decreases in progress, decrease. This noted tendency seems quite reasonable because when either  $s$  or  $m$  is increased, the total volume of data is thusly increased, and more accurate estimation results. Decided trends also exist for these two functions when the parameter values increase. The increases in the  $G_{11}$  and  $G_{22}$  values from .2 through .5 to .8 for any sample size type and level of  $r_{G_1G_2}$  are associated with very pronounced decreases in both of these functions of interest. When  $r_{G_1G_2}$  is increased from -.5 through the intermediate

levels, zero, .2, and .5 to the highest level at .8 for all sample size types and all levels of  $G_{11}$  and  $G_{22}$ , associated decreases in these two functions are noted, although these decreases are not nearly as great as the decreases associated with changes in  $s$ ,  $m$ , or  $G_{11}$  and  $G_{22}$ . The trends are partially the result of  $\Delta H$  occurring in the denominator of these functions.

The relationships of changes in these two functions,  $\frac{\Delta H - E[\Delta H']}{\Delta H}$  and  $\frac{E[(\Delta H' - \Delta H)^2]}{\Delta H}$ , to changes in  $s$ ,  $m$ ,  $G_{11}$  and  $G_{22}$  and  $r_{G_1 G_2}$  seem to be quite similar. From what can be seen in the results presented, the similarities between these two functions seem to be so great that either function is equivalent to the other for the purpose of making comparisons between different sample size types or between different levels of the parameters. This noted similarity is not really surprising when one considers the fact that the square of the second function,  $\frac{E[(\Delta H' - \Delta H)^2]}{(\Delta H)^2}$ , is equal to the square of the first function,  $\left[\frac{\Delta H - E[\Delta H']}{\Delta H}\right]^2$ , plus an added component,  $\frac{V[\Delta H']}{(\Delta H)^2}$ , due to the variation in  $\Delta H'$ , and the magnitude of this variation will be closely associated with the magnitude of the mean decrease.

Comparisons among different sample size types which have in common the same total number of offspring,  $sm$ , may be made for the first two functions. For example, for the two sample size types presented in the table where  $sm = 500$ , that is,  $s =$

50,  $m = 10$ , and  $s = 100$ ,  $m = 5$  when the  $G_{11}$  and  $G_{22}$  values are .2 or .5, the combination for  $s = 50$ ,  $m = 10$  seems preferable since the fractional decrease in progress is smaller than when  $s = 100$ ,  $m = 5$ . However, in the population types where  $G_{11} = G_{22} = .8$ , the combination for  $s = 100$ ,  $m = 5$  seems preferable. The three combinations for  $sm = 1,000$  are such that the  $s = 50$ ,  $m = 20$  combination is preferable when  $G_{11} = G_{22} = .2$ . When  $G_{11} = G_{22} = .5$ , the  $s = 100$ ,  $m = 10$  combination seems to give superior performance, and when  $G_{11} = G_{22} = .8$ , the  $s = 200$ ,  $m = 5$  groups seems superior except when  $r_{G_1G_2} = .8$ , and then the  $s = 100$ ,  $m = 10$  sample size type is slightly preferable.

For  $sm = 2,000$  the combination  $s = 100$ ,  $m = 20$  is superior when  $G_{11} = G_{22} = .2$ . When  $G_{11} = G_{22} = .5$ , the  $s = 200$ ,  $m = 10$  combinations seems to be preferable, and when  $G_{11} = G_{22} = .8$ , the  $s = 200$ ,  $m = 10$  sample size type and the  $s = 400$ ,  $m = 5$  sample size type give quite similar performance with the latter preferred when  $r_{G_1G_2}$  is small or negative and the former is preferred when  $r_{G_1G_2} = .8$ . For the combinations with  $sm = 4,000$ , when  $G_{11} = G_{22}$ , the sample size type for  $s = 200$ ,  $m = 20$  results in smaller fractional decreases in progress, while when  $G_{11} = G_{22} = .5$  or  $.8$ , the  $s = 400$ ,  $m = 10$  combination gives the smaller values for the fractional decrease in progress. For the  $sm = 8,000$  groups the  $s = 400$ ,  $m = 20$  combination gives better results than the  $s = 200$ ,  $m = 40$  for all levels of  $G_{11}$  and  $G_{22}$  and  $r_{G_1G_2}$ .

The observations on the fractional decreases in progress for different combinations of number of sires and number of offspring per sire lead to the conclusions that if  $G_{11}$  and  $G_{22}$  is large, it is desirable to have a larger number of sire groups for a fixed total number of offspring. Five offspring per sire seems in general preferable among the sample size types presented here when  $G_{11} = G_{22} = .8$ . Ten offspring per sire seem preferable when  $G_{11} = G_{22} = .5$  and 20 offspring per sire seem preferable when  $G_{11} = G_{22} = .2$ .

The above comparisons of the different sample size types were based largely upon the equational results for the two functions. As was pointed out earlier, slight discrepancies which are concluded to be due to inaccuracies of the equational results do not seem to be such as to invalidate these comparisons among the equational results. The Monte Carlo results are much too erratic to serve as a basis for such close comparisons.

The population types considered here, although they represent numerous combinations of the parameters, do not include situations where  $G_{11}$  differs from  $G_{22}$  or  $\frac{G_{11}}{P_{11}}$  differs from  $\frac{G_{22}}{P_{22}}$ . The situations considered here also do not include non-zero environmental correlations. These situations were not included in the present study because it seemed feasible to consider only a limited number of population types. One should exercise caution about extrapolating from these results to general situations because of these unexamined



situations. However, it seems feasible to speculate that the trends of the two functions measuring the decrease in progress for changes in  $s$ ,  $m$ ,  $G_{11}$  and  $G_{22}$ , and  $r_{G_1G_2}$  will be the same when  $G_{11}$  is not equal to  $G_{22}$ , and the environmental correlation is not equal to zero. Further study is needed to support this speculation.

As pointed out earlier, the values for the mean difference between the estimated progress and the expected progress for a particular calculated index are quite confusing, because discrepancies exist between the numerical values obtained from the approximate equations and their Monte Carlo estimates. The discrepancies are so large that the equational results must be rejected because it is realized that the equations are possibly inadequate and because the Monte Carlo results are correct except for sampling errors. The Monte Carlo results must be used as the basis for studying the tendency for estimates of progress to exceed or be smaller than the expected values for progress. However, the sampling errors in these Monte Carlo estimates are such as to make difficult the study of trends among these values. It seems that when  $G_{11}$ ,  $G_{22}$ , and  $r_{G_1G_2}$  are small, increases in either  $s$  or  $m$  are associated with shifts in the values for the mean difference from a positive to a negative direction. As either  $G_{11}$  and  $G_{22}$  or  $r_{G_1G_2}$  is increased in magnitude, this function tends to become negative for all levels of  $s$  and  $m$ . When  $G_{11}$  and

$G_{22}$  and  $r_{G_1G_2}$  are large, this function seems to change only very slightly as  $s$  and/or  $m$  are changed. The trends of this function associated with changes in  $s$  and  $m$  are largely obliterated by the sampling errors for those cases where  $G_{11} = G_{22} = .8$  or  $r_{G_1G_2} = .5$  or  $.8$ .

The trend, although slight, for this function to become increasingly negative as either  $s$  or  $m$  is increased seemed surprising. This trend seems to indicate that as the volume of data is increased, the "bias" will increase. Examination of the Monte Carlo estimates of the  $E[\hat{\Delta H}]$  and  $E[\Delta H']$  values (not presented here) indicates that seemingly both these values increase and approach  $\Delta H$  as either  $s$  or  $m$  increases, but the increase in  $E[\hat{\Delta H}]$  seems to be at a much lower rate. Thus the tendency for under-estimation should probably become smaller when the sample size becomes sufficiently large.

For a particular sample size type a negative change in the mean difference between  $\hat{\Delta H}$  and  $\Delta H'$  is associated with increases in either  $G_{11}$  and  $G_{22}$  or  $r_{G_1G_2}$ . These changes seem to be roughly proportional to the associated changes in the magnitude of  $\Delta H$ , the maximum attainable progress.

The numerical values which were obtained for the mean squared difference between  $\hat{\Delta H}$ , the estimate of progress, and  $\Delta H'$ , the expected progress for a particular calculated index, are such that the Monte Carlo estimates seem more reliable even though quite noticeable sampling errors occur in these

values. In the calculation of these values, it was noticed that for all population types and sample size types the magnitude of the mean squared differences was largely determined by the magnitude of the variance of  $\hat{\Delta}H$ , that is the inaccuracies of estimation are largely explained by the large variation in  $\hat{\Delta}H$  values relative to the variation in the  $\Delta H'$  values and relative to the square of the mean difference.

For a particular sample size type the magnitude of the mean squared difference increases considerably as  $G_{11}$  and  $G_{22}$  increase. Its magnitude also increases as  $r_{G_1G_2}$  increases. As might be expected, the changes in magnitude of the mean squared difference for a particular sample size seems to be roughly, though not exactly, proportional to the changes in the square of the  $\Delta H$  values for that population type. For a particular population type, this function tends to decrease as either  $s$  or  $m$  increases. This trend is considerably stronger in those population types where  $G_{11}$  and  $G_{22}$  and  $r_{G_1G_2}$  are small, and the trend is hardly detectable when  $G_{11}$  and  $G_{22}$  and  $r_{G_1G_2}$  are large.

The presented numerical values and the trends among them may be used to provide an indication of the probable effectiveness of a particular set of data arranged in paternal-half-sib groups for the purpose of index construction. However, the use of these tables is rendered quite difficult because the effectiveness of an estimation procedure is so largely

affected by the true parameters. In other words, the ability to estimate parameters is dependent upon what these parameters are. However, in some cases the experimenter will have a rough idea of the magnitude of the parameters he is estimating prior to the collection and analysis of his data.

The question most likely to be asked by an experimenter is, "What volume of data shall be sufficient for the construction of an adequate selection index?" Besides depending on the magnitude of the true parameters, the answer to this question also depends upon the definition of an adequate selection index. This definition will be related to the two following factors: (1) the nearness of the expected gain from this index to the maximum possible gain and (2) the accuracy of predicting the gain from selection upon the particular calculated index. The levels necessary for the satisfaction of these two criteria of adequacy cannot be generally defined, and any conclusions about the effectiveness of an estimation procedure will depend on these factors. Since any answers to the question posed above will be so dependent upon definitions and personal opinion, the answer for a particular situation shall be left for the reader to determine after reasonable study and comparison of the numerical values and trends in the tables with the known facts of the pertinent situation.

However, it seems from these tables that a "satisfactory"

closeness of the distribution of gains to the maximum attainable gain will likely be achieved from a volume of data much less than the volume necessary for "reasonable" accuracy in predicting genetic gains. The relative importance of the two attributes of a particular estimation procedure discussed above will vary. The ability to make near maximum gains will always be of great importance. However, the ability to predict this gain will be of more importance in some situations than in others. For example, when other breeding schemes such as hybridization, cross mating systems, and numerous others are also being considered, the ability to accurately predict the gain will be of definite importance for comparisons with other breeding schemes. In these latter cases it seems, from the tables presented here, that much more voluminous data will be necessary in the index construction.

The use of these tables is restricted because the population types which were studied here do not satisfactorily represent all the population types that might be of interest. The restriction of the population types to those with phenotypic variances of unity does not seem detrimental to the over-all interpretation of the results. Phenotypic variances of unity could be developed from any population type by the appropriate coding of the phenotypic values. Thus the above comparisons among different levels of  $G_{11}$  and  $G_{22}$  are actually comparisons among levels of  $\frac{G_{11}}{P_{11}}$  and  $\frac{G_{22}}{P_{22}}$ , the heritability

values for the two traits of interest.

The population types considered here include only situations where  $G_{11}$  is equal to  $G_{22}$ , that is, only situations where the heritabilities of the two traits of interest are the same. Also, the economic importance for the two traits was considered to be the same in the population types studied. These population types do not include any situations in which the environmental correlations,  $\frac{P_{12} - G_{12}}{(P_{11} - G_{11})^{\frac{1}{2}}(P_{22} - G_{22})^{\frac{1}{2}}}$ , between the two traits are non-zero.

All these situations, which are not included in the population types studied here, make any general conclusions ill advised. More than likely the trends noted and discussed here will also prevail in the other possible situations, but further study is necessary in order to establish this speculation.

Selection indexes for more than two traits are quite common in the applied situations. The principles of construction are the same, and it thus seems that the trends which occur for the two trait selection indexes will have counterparts in the indexes for three or more traits.

In the situations evaluated here the  $a_i$  values, the economic weights, were treated as constants. However, in many applied situations the  $a_i$  values are not known and some sort of estimation procedure is involved in determining them. The inaccuracies of these estimation procedures will probably

also influence the realized gain from selection.

It is hoped that the present work may be extended in order that any general conclusions which seem plausible on the basis of the present work may be definitely and firmly established as being generally true for all realistic situations. This combined body of knowledge would then give deeper understanding of index selection and its effectiveness and reliability as a tool for animal and plant breeding. As was pointed out in an earlier section, the theory of selection index construction has found widespread application in several commercial species. It is believed that the values presented in the foregoing tables along with the observed trends among them will lead to a fuller and more complete understanding of this breeding technique in spite of the limitations of the results presented. Certainly further work should remove some of the limitations of the results which are complete at present.

The estimation procedure involving paternal-half-sib analyses of variance and covariance was the primary procedure considered. Numerous other procedures have been used and will be used in the future for index construction. It is hoped that future work can include some or all of these estimation procedures. Such results would facilitate comparisons among the different estimation procedures which might in time lead to more effective experimental design.

It should also be realized that the results presented were obtained in such a manner that there was complete satisfaction of all the models and assumptions involved in the selection index theory and in the estimation procedure. Unfortunately, this may not always be true of any real set of data. So the validity of the use of the results for making inferences about realistic situations is dependent upon how nearly the data in the realistic situations satisfy the assumptions which are necessary. The assumptions which are least likely to be satisfied are those involved in the model,  $X_i = G_i + E_i$ , which ignores the effects of dominance and epistasis. The inappropriateness of this model is especially pronounced in the estimation procedure studied here since the sire components of variance and covariance are biased by contributions from factors other than the additive genetic effects of the genes. The inadequacies of this model are the causes of many of the weaknesses of quantitative genetic theory at present. However, a more exact model would not have comparable simplicity, and therefore the use of the simpler model will probably be continued for some time to come. In addition, the assumptions of normality of distributions might not be completely satisfied at times.

Despite these limitations, the results presented here give indications of trends that were not known or at least the magnitude of the trends was not known prior to this study.



With successful elaboration upon this study the effectiveness and accuracy of index selection should be more fully understood.

## SUMMARY

The influence of errors of parameter estimation upon index selection has been studied using two techniques. The first of these was to develop approximate equations for certain functions of interest, while the second was the Monte Carlo simulation of the sampling distribution of the parameter estimates and a study of the resulting progress and estimates of progress. Unfortunately, the two methods of study did not give equivalent results, and due to the realized possibility of inaccuracies in the approximate equations, the Monte Carlo results were taken to be more realistic in spite of quite large sampling errors in these results.

Of primary interest in this study were: (1) the closeness of the distribution of realized gain from a particular calculated index to the maximum possible gain from selection and (2) the accuracy of the usual estimates of progress as estimators of the true expected gains for a particular calculated index. In all situations the gains were considered to be those which would result when selection was carried out in an infinite population, and the further variation due to the finiteness of the population in which selection was carried out was excluded from consideration in this study.

Even with the large sampling errors encountered in the Monte Carlo results certain trends are apparent from this

study. When the heritability values for both of the two traits considered in an index are increased, the fractional decrease in progress becomes smaller. Similar trends for the fractional decrease in progress are noted when the genetic correlation between the two traits is increased, when the number of paternal-half-sib groups used for estimation is increased, and when the number of offspring per sire group is increased.

The mean difference between the estimates of genetic gain and the expected genetic gain for particular calculated indexes tends to be negative for most combinations of the true parameters and for most combinations of number of sires and number of offsprings per sire. This indicates that there is a tendency for under-estimation of the progress from index selection. A tendency is noted for the mean squared difference between the estimated progress and the expected progress for calculated indexes to increase as the heritabilities for the two traits increase and as the genetic correlation for the two traits increases. For a particular combination of the true parameters this latter function, which is a measure of the accuracy of predicting genetic progress from index selection, tends to decrease slightly as either the number of sires or the number of offspring per sire increases.

Generalizations upon these observed trends do not seem

feasible or desirable at the present time because the combinations of true parameters which were studied here do not include situations where the environmental correlation between the two traits is non-zero. However, it is hoped that the present work with the desired future extensions will lead to a deeper understanding of index selection and its relationship to estimation procedures.

## BIBLIOGRAPHY

- Aitken, A. C. 1954. Determinants and matrices. 8th ed. Edinburgh, Scotland, Oliver and Boyd.
- Anderson, R. L. and T. A. Bancroft. 1952. Statistical theory in research. New York, N. Y., McGraw-Hill Book Co., Inc.
- Barker, J. S. F. 1958a. Simulation of genetic systems by automatic digital computers. III. Selection between alleles at an autosomal locus. Aust. J. Biol. Sci. 11:603-612.
- Barker, J. S. F. 1958b. Simulation of genetic systems by automatic digital computers. IV. Selection between alleles at a sex-linked locus. Aust. J. Biol. Sci. 11:613-625.
- Bartlett, M. S. 1939. The standard errors of discriminant function coefficients. J. Roy. Stat. Soc. (Suppl.) 1939: 169-173.
- Browne, E. T. 1958. Introduction to the theory of determinants and matrices. Chapel Hill, N. C., The University of North Carolina Press.
- Cochran, W. G. 1951. Improvement by means of selection. Proc. of the Second Berkeley Sum. on Math, Stat, and Probability. 1951: 449-470.
- Crittenden, L. B. 1958. Factors affecting the efficiency of selection for hatchability. Unpublished Ph. D. Thesis. Lafayette, Ind., Library, Purdue University. (Original not available for examination; abstracted in Diss. Absts. 18: 1925. 1958.)
- Crump, S. L. 1947. The estimation of components of variance in multiple classifications. Unpublished Ph. D. Thesis. Ames, Iowa, Library, Iowa State University of Science and Technology.
- Ercanbrack, S. K. 1952. Selection indexes for range Rambouillet, Columbia and Targhee lambs. Unpublished Ph. D. Thesis. Ames, Iowa, Library, Iowa State University of Science and Technology.

- Evans, Lee. 1956. Estimation of genetic parameters necessary for constructing selection indexes for beef cattle. Unpublished Ph. D. Thesis. Urbana, Ill., Library, University of Illinois. (Original not available for examination; abstracted in Diss. Absts. 16: 837. 1956.)
- Farnsworth, G. M. 1956. Estimates of genetic parameters influencing blood spots and other economic traits of the fowl. Unpublished Ph. D. Thesis. Ames, Iowa, Library, Iowa State University of Science and Technology.
- Felts, L. C. 1958. Construction of a ewe selection index for use under farm flock conditions. Unpublished Ph. D. Thesis. Madison, Wis., Library, University of Wisconsin. (Original not available for examination; abstracted in Diss. Absts. 18: 1925. 1958.)
- Fisher, R. A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. Trans. Roy. Soc. Edinburgh. 52, Pt. 2: 399-433. 1918.
- Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. Ann. Eugen. 7: 179-189.
- Fraser, A. S. 1957a. Simulation of genetic systems by automatic digital computers. I. Introduction. Aust. J. Biol. Sci. 10: 484-491.
- Fraser, A. S. 1957b. Simulation of genetic systems by automatic digital computers. II. Effects of linkage on rates of advance under selection. Aust. J. Biol. Sci. 10: 492-499.
- Ghostléy, J. E. 1955. Genetic relationships between early growth, mature body weight and egg size in the fowl. Unpublished Ph. D. Thesis. Ames, Iowa, Library, Iowa State University of Science and Technology.
- Givens, C. S., Jr., R. C. Carter and J. A. Gaines. 1960. Selection indexes for weanling traits in spring lambs. J. Anim. Sci. 19: 134-139.
- Hanson, W. D. and H. W. Johnson. 1957. Methods of calculating and evaluating a general selection index obtained by pooling information from two or more experiments. Genetics. 42: 421-432.
- Harvey, W. R. and J. L. Lush. 1952. Genetic correlation between type and production in Jersey cattle. J. Dairy Sci. 35: 199-213.

- Hazel, L. N. 1943. The genetic basis for constructing selection indexes. *Genetics*. 28: 476-490.
- Hazel, L. N. and J. L. Lush. 1942. The efficiency of three methods of selection. *J. Heredity*. 33: 393-399.
- Hazel, L. N. and C. E. Terrill. 1945. Heritability of weaning weight and staple length in range Rambouillet lambs. *J. Anim. Sci.* 4: 347-358.
- Hazel, L. N. and C. E. Terrill. 1946. The construction and use of a selection index for range Rambouillet lambs. (Abstract) *J. Anim. Sci.* 5: 412.
- Henderson, C. R. 1951. [Title not given]. (Mimeo.) Ithaca, N. Y., Dept. of Animal Husbandry, Cornell University, (Original not available for examination; cited in Tabler, K. A. and R. W. Touchberry. 1955. Selection indices based on milk and fat yield, fat percent, and type classification. *J. Dairy Sci.* 38: 1155-1163.)
- Hogsett, M. L. and A. W. Nordskog. 1958. Genetic-economic value in selecting for egg production rate, body weight, and egg weight. *Poultry Sci.* 37: 1404-1419.
- I.B.M. reference manual; random number generation and testing. 1959. New York, N. Y., International Business Machines Corporation.
- Karam, H. A. 1959. Selecting yearling Rahmani sheep. *J. Anim. Sci.* 18: 1452-1459.
- Karam, H. A., A. B. Chapman and A. L. Pope. 1953. Selecting lambs under farm flock conditions. *J. Anim. Sci.* 12: 148-163.
- Kempthorne, Oscar. 1957. An introduction to genetic statistics. New York, N. Y., John Wiley and Sons, Inc.
- Kempthorne, Oscar and A. W. Nordskog. 1959. Restricted selection indices. *Biometrics*. 15: 10-19.
- Kendall, M. G. and A. Stuart. 1958. The advanced theory of statistics. Volume I. Distribution theory. New York, N. Y., Hafner Publishing Co.

- Krueger, W. F., G. E. Dickerson, Q. B. Kinder and H. L. Kempster. 1952. The genetic and environmental relationship of total egg production to its components and to body weight in the domestic fowl. (Abstract) Poultry Sci. 31: 922-923.
- Latter, B. H. D. and Alan Robertson. 1960. Experimental design in the estimation of heritability by regression methods. Biometrics. 16: 348-353.
- Legates, J. E. and J. L. Lush. 1950. A selection index for fat production utilizing the fat yields of the cow and her relatives.
- Lerner, I. M., V. S. Asmundson and Dorothy Cruden. 1947. The improvement of New Hampshire fryers. Poultry Sci. 26: 515-524.
- Lindholm, H. B. and H. H. Stonaker. 1957. Economic importance of traits and selection indexes for beef cattle. J. Anim. Sci. 16: 998-1006.
- Lush, J. L. 1940. Intrasire correlations or regressions of offspring on dam as a method of estimating heritability of characteristics. Am. Soc. Anim. Prod. Proc. 33: 293-301. 1940.
- Lush, J. L. 1948. The genetics of populations. (Mimeo.) Ames, Iowa, Department of Animal Husbandry, Iowa State University of Science and Technology.
- Malécot, G. 1948. Les Mathématiques de l'hérédité. Paris, France, Masson et Cie.
- Manning, H. L. 1956. Yield improvement from a selection index technique with cotton. Heredity. 10: 303-322.
- Merritt, E. S. 1956. Genetic and environmental effects on body measures in broiler chickens. Unpublished Ph. D. Thesis. Ames, Iowa, Library, Iowa State University of Science and Technology.
- Morley, F. H. W. 1950. Selection for economic characters in Merino sheep. Unpublished Ph. D. Thesis. Ames, Iowa, Iowa State University of Science and Technology.
- Nagai, J., M. Yoshida and M. Naito. 1955. Individual selection of mice by the selection index method. Jap. J. Zootech. Sci. 26(Suppl. 2): 3. (Original not available for examination; abstracted in Anim. Breed. Abst. 26: 347. 1958.)



- Nanda, D. N. 1949a. The standard errors of discriminant function coefficients in plant-breeding experiments. J. Roy. Stat. Soc. Ser. B, 11: 283-290.
- Nanda, D. N. 1949b. Efficiency of the application of discriminant function in plant selection. J. Ind. Soc. Agr. Stat., 2: 8-19.
- Nordskog, A. W. and J. F. Hill. 1958. Correlation between egg production and adult viability in hybrid flocks. Poultry Sci. 37: 1265-1273.
- Osborne, R. and W. S. B. Patterson. 1952. On the sampling variance of heritability estimates derived from variance analyses. Proc. Roy. Soc. Edinburg. 64: 456-461.
- Panse, V. G. 1946. An application of the discriminant function for selection in poultry. J. Genetics. 47: 242-248.
- Rae, A. L. 1950. Genetic variation and covariation in productive characters of New Zealand Romney Marsh sheep. Unpublished Ph. D. Thesis. Ames, Iowa, Iowa State University of Science and Technology Library.
- Reeve, E. C. R. 1955. The variance of the genetic correlation coefficient. Biometrics. 11: 357-374.
- Robertson, Alan. 1959a. The sampling variance of the genetic correlation coefficient. Biometrics. 15: 469-485.
- Robertson, Alan. 1959b. Experimental design in the evaluation of genetic parameters. Biometrics. 15: 219-226.
- Sidwell, G. M. 1954. A selection index for Navajo crossbred range lambs. Unpublished Ph. D. Thesis. Ames, Iowa, Library, Iowa State University of Science and Technology.
- Smith, H. F. 1936. A discriminant function for plant selection. Ann. Eugen. 7: 240-250.
- Sutherland, T. M. 1958. An index for selecting hogs using data from a testing station. Unpublished Ph. D. Thesis. Ames, Iowa, Library, Iowa State University of Science and Technology.
- Tabler, K. A. and R. W. Touchberry. 1955. Selection indices based on milk and fat yield, fat percent, and type classification. J. Dairy Sci. 38: 1155-1163.

- Tabler, K. A. and R. W. Touchberry. 1959. Selection indices for milk and fat yield of Holstein-Friesian dairy cattle. *J. Dairy Sci.* 42: 123-126.
- Tallis, G. M. 1959. Sampling errors of genetic correlation coefficients calculated from analyses of variance and covariance. *Aust. J. Stat.* 1: 35-43.
- Tallis, G. M. 1960. The sampling errors of estimated genetic regression coefficients and the errors of predicted genetic gains. *Aust. J. Stat.* 2: 66-77.
- Tallis, G. M. and E. W. Klosterman. 1959. Efficient estimates of heritability from paternal half-sib correlations. *J. Anim. Sci.* 18: 622-628.
- Van Vleck, L. D., S. R. Searle and C. R. Henderson. 1960. The number of daughter-dam pairs needed for estimating heritability. *J. Anim. Sci.* 19: 916-920.
- Wadell, L. H. and G. V. O'Brien. 1960. Empirical sampling in quantitative genetic research. Paper presented at Data Processing Oriented Advanced Statistical Methods Seminar, Poughkeepsie, N. Y. May 2-4, 1960. (Mimeo.) Ithaca, N. Y., Dept. of Animal Husbandry, Cornell University.
- Warren, W. M. and G. E. Dickerson. 1952. Components of performance in selecting for heterosis in swine. *Missouri Agr. Expt. Sta. Res. Bull.* 511.
- Winters, L. M., D. L. Dailey, O. M. Kiser, P. S. Jordan, R. E. Hodgson and W. W. Green. 1946. Factors affecting productivity in breeding sheep. *Minnesota Agr. Expt. Sta. Tech. Bul.* 174.
- Wright, S. 1921. Systems of mating. I. The biometric relations between parent and offspring. *Genetics* 6: 111-123.
- Wyatt, A. J. 1954. Genetic variation and covariation in egg production and other economic traits in chickens. *Poultry Sci.* 33: 1266-1274.
- Yamada, Y. 1958. Heritability and genetic correlations in economic characters in chickens. *Jap. J. Genetics.* 33: 13-22.
- Young, C. W., J. E. Legates and J. G. Lecce. 1960. Genetic and phenotypic relationships between clinical mastitis, laboratory criteria, and udder height. *J. Dairy Sci.* 43: 54-62.

Young, S. S. Y. 1959. The effects of sex differences in heritability and repeatability on Merino breeding plans. Unpublished Ph. D. Thesis. Sydney, N. S. W., Australia, Library, University of Sydney.

## ACKNOWLEDGEMENTS

The author wishes to express his sincere appreciation to Professors L. N. Hazel, O. Kempthorne, and J. L. Lush for their stimulation and encouragement during the course of this study and their constructive suggestions during the preparation of this manuscript.

The author is deeply indebted to Mr. Howard Jespersen of the Statistical Laboratory who spent long and tedious hours on the development and operation of the IBM 650 program for the Monte Carlo phase of this study.

Words can hardly be indicative of the gratefulness felt by the author for the moral support, assistance, and encouragement rendered by his beloved wife, Eugenia, throughout this study and throughout the author's graduate work.

## APPENDIX A

## Glossary of Symbols

	(Pages on which symbols are defined)
$a$	pages 12 and 13
$a_i$	page 3
$b$	pages 12 and 13
$B_{HI}$	page 5
$B_i$	page 10
$b_i$	pages 6 and 8
$\wedge$	page 24
$Cov_c( )$	pages 40 and 46
$Cov[ ]$	page 46
$\Delta g_w$	page 47
$\Delta H$	pages 43 and 58
$\Delta H'$	pages 43 and 58
$\hat{\Delta H}$	pages 53 and 58
$\Delta p$	page 47
$E_i$ (variety selection)	page 7
$E_i$ (individual selection)	page 13
$E[ ]$	page 46
$G$	pages 12 and 13
$G_i$ (variety selection)	page 3
$G_i$ (individual selection)	page 13
$G_{ij}$ (variety selection)	pages 7 and 8

$G_{ij}$ (individual selection)	page 14
$G_{iW}$	page 16
$G_W$	page 17
$H$	page 3
$I$	page 4
$i$	page 5
$\bar{I}_S$	page 5
$\bar{I}_S$	page 5
$M$	page 32
$m_K$	page 32
$m_O$	pages 32 and 33
$\mu_H$	page 5
$\mu_I$	page 5
$P$	page 12
$P_{ij}$	page 8
$P_W$	page 17
$r_{E_1 E_2}$	page 85
$r_{G_1 G_2}$	page 85
$R_{IH}$	page 10
$r_{X_i X_t}$	page 10
$r_{X_t H}$	page 10
$s$	page 32
$s_{ij}$	page 32

$\sigma_{ije}$	pages 32 and 33
$\sigma_{ijs}$	pages 32 and 33
$t_{ij}$	page 32
$v_c( )$	pages 40 and 46
$v[ ]$	page 46
$w_{ij}$	page 32
$x_i$	pages 6 and 7